INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

18880

I A R.I. 6
MGIPC—S4—10 AR—21-6 49—1 000

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL CLXV

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W. 1

27 April 1938

18880

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOL. CLXV

No. A 920 18 March 1938

PAGE

Critical phenomena in gases. II. Vapour pressures and boiling points. By J. E. Lennard-Jones, F.R.S. and A. F. Devonshire 1

The nature of the penetrating component of cosmic rays. By P. M. S. Blackett, F.R.S. 11

The continuous absorption spectra of alkyl iodides and alkyl bromides and their quantal interpretation. By D. Porret and C. F. Goodeve 31

The infra-red absorption spectrum of methylene chloride. By C. Corin and G. B. B. M. Sutherland 43

On the velocity and temperature distributions in the turbulent wake behind a heated body of revolution. By S. Tomotika 53

Application of the modified vorticity transport theory to the turbulent spreading of a jet of air. By S. Tomotika 65

An experimental determination of the spectrum of turbulence. By L. F. G. Simmons and C. Salter 73

Comparison of wave-functions for HeH^{++} and HeH^{+}. By C. A. Coulson and W. E. Duncanson 90

The production of gamma-rays by neutrons. By E. H. S. Burhop, R. D. Hill and A. A. Townsend 116

The formation of mercury molecules. II. By F. L. Arnot and Marjorie B. M'Ewen 133

Oscillography of adsorption phenomena. III. Rates of deposition of oxygen upon tungsten. By M. C. Johnson and A. F. Henson 148

No. A 921 5 April 1938

Significance tests when several degrees of freedom arise simultaneously. By H. Jeffreys, F.R.S. 161

✓ A new basis for cosmology. By P. A. M. Dirac, F.R.S. 199

The scattering of cosmic ray particles in metal plates. By P. M. S. Blackett, F.R.S. and J. G. Wilson 209

PAGE

On the motion of a fluid heated from below By R J Schmidt and O. A. Saunders. (Plates 1, 2) 216

Identification and measurement of helium formed in beryllium by γ-rays By E Gluckauf and F A Paneth 229

On the occurrence of helium in beryls By J W J Fay, E Gluckauf and F A. Paneth 238

On the neutrino theory of light By M H. L Pryce 247

The absorption spectra of sulphur dioxide and carbon disulphide in the vacuum ultra-violet By W C Price and (Miss) D M Simpson. (Plates 3, 4) 272

A suggestion for unifying quantum theory and relativity By M Born 291

Hyperfine structure, Zeeman effect and isotope shift in the resonance lines of potassium. By D A Jackson and H Kuhn (Plate 5) 303

No A 922 14 April 1938

On the equations of electromagnetism I Identifications By E. A Milne, F R S 313

On the equations of electromagnetism II Field Theory By E A Milne, F R S 333

The crystalline structure of steel at fracture By H J. Gough, F R S. and W A Wood (Plates 6–9) 358

Collective electron ferromagnetism By Edmund C Stoner, F R S 372

Resonance in crystal beams of sodium-ammonium seignette salt By W Mandell 414

The alkaline permanganate oxidation of organic substances selected for their bearing upon the chemical constitution of coal By R B Randall, M Benger and C. M. Groocock 432

No A 923. 27 April 1938

Photo-electric measurements of the seasonal variations in daylight around 0 41μ, from 1930 to 1937. By W R G Atkins, F R S 453

Quantitative spectrographic analysis of biological material. III A method for the determination of sodium and potassium in glandular secretions. By J S Foster, F R S, G O Langstroth and D R. McRae 465

PAGE
Thermal conduction in hydrogen-deuterium mixtures. By C. T. Archer . 474
The theory of pressure-ionization and its applications By D S. Kothari 486
The influence of wall oscillations, wall rotation, and entry eddies, on the breakdown of laminar flow in an annular pipe By A Fage 501
Uranium Z and the problem of nuclear isomerism By N Feather and E Bretscher 530
The adsorption of vapours at plane surfaces of mica Part I By D H. Bangham and S Mosallam 552
The influence of rate of deformation on the tensile test with special reference to the yield point in iron and steel By C F. Elam 568
Index . . . 593

Critical phenomena in gases
II. Vapour pressures and boiling points

By J. E. Lennard-Jones, F.R.S. and A. F. Devonshire
The University Chemical Laboratory, Cambridge

(*Received* 24 *November* 1937)

1. Introduction

The object of the first paper on "Critical Phenomena in Gases" (referred to in this paper as Paper I) was to develop a simple method of dealing with dense gases and to calculate critical temperatures in terms of atomic fields of force. Each atom in a dense gas was pictured as caged for most of its time by a cluster of neighbours, equal in number to those which surround it in the solid (and presumably also in the liquid) phase The model was intended to provide a general average of the potential field in which any one atom moved by replacing its varying environment by an arrangement of neighbours which could be regarded as typical. This arrangement was taken to be the one in which the neighbours were situated at their mean positions as determined by the density of the gas The potential energy of any one atom could thus be expressed as a function of the volume of the gas—a step which is probably the crucial one in a theory of critical phenomena. This point of view brings the process of condensation within the category of those described by Fowler (1936) as *co-operative* phenomena In passing we may observe that the derivation of van der Waals' equation provides a particular example of this method, for in it the potential energy of each atom is assumed to be inversely proportional to the volume. The present theory goes beyond this simple model, for the potential energy of an atom is considered to be not only a function of volume but also a function of the position of the atom relative to its neighbours. The probability of finding an atom in any assigned position can be calculated by statistical means and its average potential energy and its available free volume easily deduced. The equation of state can then be deduced by thermodynamic methods, as has been pointed out elsewhere (Lennard-Jones 1937).

The success of this method in calculating critical temperatures has encouraged an attempt to push the theory a stage further so as to give the boiling points of gases. At the boiling-point the conditions in the gas are very different from those at the critical point. The density is much less and the perfect gas laws can be assumed to hold But in the liquid we may suppose that the conditions are similar to those of a dense gas except for the closer relation of neighbouring atoms In this paper therefore we use the same methods for the liquid which we adopted for the dense gas in the first paper and then find a relation between the temperature and vapour pressure. This enables us not only to calculate boiling-points in terms of atomic force fields but also to determine the change of entropy on evaporization and thus to give a theoretical justification of Trouton's rule

While we have approached the problem of a liquid and its vapour as a natural extension of the theory of dense gases, we find that considerable work has already been done on the theory of liquids, which bears some resemblance to that developed here T S Wheeler (1934–6) has used the force fields previously given by Lennard-Jones as a basis for a theory of liquids He supposes each molecule to keep clear round it a spherical volume by its thermal motion and to vibrate within it like a linear oscillator. He is thus able to obtain many properties of a liquid by simple kinetic theory arguments in terms of the specific volume and the constants of the force fields The theory does not determine the specific volume, as a self-contained theory should do, but by giving to this quantity its experimental value he is able to get satisfactory agreement between theory and experiment for a number of physical properties

Recently Eyring and his collaborators (1936–7) have developed a theory of the liquid state which seeks not to obtain all the properties of a liquid in terms of interatomic forces but rather to correlate the different properties with one another Eyring has introduced a new concept which is extremely valuable He has given reasons for supposing that in a liquid there are a number of "holes" of atomic size, the number of which can be estimated by simple thermodynamic arguments. In terms of these the phenomena of diffusion and viscosity can be dealt with quantitatively. In a further paper (Newton and Eyring 1937) the vapour pressure has been obtained in terms of the coefficient of expansion and the specific heats and other properties have been successfully correlated (Eyring and others 1937).

An attempt to develop an exact theory of condensing systems has recently been made by Mayer (1937). This work must be regarded as an important advance in the subject, but though the vapour pressure of a liquid is expressed in terms of the interatomic fields, the equations are too complicated to admit of more than a very approximate calculation In order to make applications to particular gases and liquids, it will probably be necessary for some time to try to find methods, such as that given in this paper, which permit of easy numerical computation.

2 CALCULATION OF VAPOUR PRESSURES

In calculating the partition function and hence the free energy of a dense gas in Paper I, we supposed that each atom could be regarded as confined by its immediate neighbours to a cell or cage and that its migration from one environment to another could be regarded as a relatively infrequent event In conformity with this, we supposed the partition function of the assembly to be equal to the product of the partition functions of the individual molecules, but the fact that migration may occur, though slowly, implies that the statistics should be like those of a dilute gas rather than that of a solid We discussed in Paper I the partition function for a perfect gas by these two methods and showed the ability to change places results in a factor of e^N in the partition function for the assembly (corresponding to a factor $N^N/N^N!$). As we stated in that paper, none of the results given there (for example, the equation of state and critical constants) are affected by this factor, but in extending the results in this paper to vapour pressures this factor must be considered more closely. The contribution of mobility to the energy of the assembly will be small and may safely be neglected, but it is probably more accurate to take into account the permutation of the atoms and to include the factor just referred to. This means that the whole of the "free volume" of the assembly and not just that of each cell is accessible to every atom and the usual factor of $N!$ must be introduced to allow for permutations. This method has already been used by Eyring and Hirschfelder in a paper just published (1937)

In Paper I we supposed the fields of the atoms to be spherically symmetrical and the potential to be of the special form

$$\phi(r) = Ar^{-n} - Br^{-m}. \tag{1}$$

Now atoms obeying a law of force of this type would crystallize in the form of a face-centred cubic (Lennard-Jones and Ingham 1925) and the number of nearest neighbours of each atom would be 12 We accordingly assume that in the liquid and dense gas phases there is a tendency to this structure and that the average field in which any one moves can be represented approximately by the effect of the 12 nearest neighbours in their mean positions. Clearly it will be necessary in taking the theory to further approximations to consider the motion of the atoms about those mean positions and to represent this by probability patterns as is done for electrons in atoms, in fact, what is ultimately required is a method analogous to that of self-consistent fields for finding electronic distributions.

In this paper we shall suppose the methods used for a dense gas are applicable also to a liquid. The partition function, obtained in Paper I, equation (48), can be written in the form

$$\frac{1}{N_l}\log F(T) = -\frac{\Phi_0}{N_l kT} + \tfrac{3}{2}\log(2\pi m\, kT/h^2) + \log(v_l\chi/N_l), \tag{2}$$

where $$\Phi_0 = -N_l A\{(1\cdot 2)\,(v_0/v_l)^2 - (0\cdot 5)\,(v_0/v_l)^4\}, \tag{3}$$

and is the mutual potential energy of N_l atoms, when arranged in a face-centred cubic lattice at their mean positions. This is easily obtained from (1) by adopting the summations given by Lennard-Jones and Ingham (1925) and by altering the notation so as to express the energy in terms of volume, thus,

$$A = c\phi_0, \quad v_0 = N_l r_0^3/\sqrt{2}, \tag{4}$$

where ϕ_0 is the absolute value of the minimum of the potential energy of two atoms under the field (1), c the number of nearest neighbours ($= 12$) and r_0 their distance apart in the equilibrium configuration. N_l is the number of atoms in the assembly, considered to be in the liquid phase, v_l the volume occupied by these atoms, while χ is the fraction of the total volume which may be regarded as available to an atom, in fact,

$$\chi = (1/v^*)\int \exp[-\psi(r)/kT]\,d\tau, \tag{5}$$

where the integral is taken over a unit cell of volume v^* (the specific volume) and $\psi(r)$ is the potential energy of the atom within its cell referred to that at its centre as zero For the particular law of force defined by (1) χ is given by

$$\chi = 2\pi\sqrt{2}\int_0^{\frac{1}{4}} y^{\frac{1}{2}} \exp\left[\frac{A}{kT}\{-(v_0/v_l)^4\, l(y) + 2(v_0/v_l)^2\, m(y)\}\right] dy \tag{6}$$

(cf. equation (49), Paper I, using the relation $\chi = 2\pi\sqrt{2g}$)

We may conveniently define a free volume v_f by the relation

$$v_f = v_l\chi. \tag{7}$$

We note that it is a function of temperature and volume. The free volume per atom we shall denote by v_f^*

As explained above we now modify the partition function from (2) to

$$(1/N_l)\log F(T) = -\Phi_0/N_l kT + \tfrac{3}{2}\log(2\pi m\, kT/h^2) + \log v_f^* + 1. \tag{8}$$

The vapour pressure can be calculated from the well-known thermodynamical formula

$$T\left(\frac{\partial p}{\partial T}\right) - p = \left(\frac{\partial U}{\partial v}\right), \tag{9}$$

where U is the internal energy of the substance in question. Now p remains constant and equal to the vapour pressure while the substance evaporates at constant temperature, so that if we integrate (9) at constant temperature from $v = v_l^*$ (the specific volume of the liquid), to $v = v_g^*$ (the specific volume of the gas), we obtain the equation

$$\left\{T\left(\frac{\partial p}{\partial T}\right)_v - p\right\}(v_g^* - v_l^*) = U_g^* - U_l^*, \tag{10}$$

where U_g^* is the internal energy per molecule of the gas, U_l^* the internal energy per molecule of the liquid, and p the vapour pressure. If the vapour pressure is small (of the order of an atmosphere or less) we can neglect v_l^* in comparison with v_g^* and treat the vapour as a perfect gas. Then we have

$$\left\{T\left(\frac{\partial p}{\partial T}\right)_v - p\right\}(kT/p) = \tfrac{3}{2}kT - U_l^*. \tag{11}$$

Now $$U_l^* = \frac{kT^2}{N_l}\frac{\partial}{\partial T}\log F(T) = \frac{\Phi_0}{N_l} + \frac{kT^2}{\chi}\frac{\partial \chi}{\partial T} + \tfrac{3}{2}kT. \tag{12}$$

so that $$\left\{T\left(\frac{\partial p}{\partial T}\right)_v - p\right\}\frac{kT}{p} = -\frac{\Phi_0}{N_l} - \frac{kT^2}{\chi}\frac{\partial \chi}{\partial T} \tag{13}$$

If we divide by kT^2 and integrate we have

$$\log p - \log T = \Phi_0/N_l kT - \log\chi + \log f(v),$$

so that $$p = T\chi^{-1}f(v)\exp\{\Phi_0/N_l kT\}, \tag{14}$$

where $f(v)$ is an arbitrary function of v. Since χ is a pure number, it would appear from dimensions that $f(v)/k$ must be equal to the inverse of a volume.

Another method of calculating the vapour pressure, which has the advantage of giving the equation without any arbitrary function, is to use the thermodynamical equations for the condition that two phases should be in equilibrium in the form

$$\left(\frac{\partial A_l}{\partial v_l}\right)_{N_l} = \left(\frac{\partial A_g}{\partial v_g}\right)_{N_g}, \tag{15}$$

$$\left(\frac{\partial A_l}{\partial N_l}\right)_{v_l} = \left(\frac{\partial A_g}{\partial N_g}\right)_{v_g}, \tag{16}$$

where A is the Helmholtz free energy or work function, which, for the liquid, is given by the equation (cf. equation (48), Paper I)

$$A_l = -kT\log F_l(T) = \Phi_0 - N_l kT\{\log(v_l\chi/N_l) + 1\} - \tfrac{3}{2}N_l kT\log(2\pi m kT/h^2), \tag{17}$$

and for the vapour, assuming it to be a perfect gas, by

$$A_g = -N_g kT\{\log(v_g/N_g)+1\} - \tfrac{3}{2}N_g kT\log(2\pi mkT/h^2). \tag{18}$$

Equation (15) is the condition that the pressure on the liquid should be the vapour pressure, and is

$$-p_g = \frac{-N_g kT}{v_g} = -p_l = \frac{-N_l kT}{\chi}\frac{\partial\chi}{\partial v_l} - \frac{N_l kT}{v_l} + \frac{\partial\Phi_0}{\partial v_l},$$

or

$$p_g = \frac{N_l kT}{v_l} + \frac{N_l kT}{\chi}\frac{\partial\chi}{\partial v_l} - \frac{\partial\Phi_0}{\partial v_l}. \tag{19}$$

Equation (16) gives the relation*

$$-kT\log(v_g/N_g) = \Phi_0/N_l - kT\log\chi - kT\log(v_l/N_l) - \frac{v_l}{N_l}\left\{\frac{\partial\Phi_0}{\partial v_l} - N_l kT\frac{1}{\chi}\frac{\partial\chi}{\partial v_l}\right\}$$

$$= \Phi_0/N_l - kT\log\chi - kT\log(v_l/N_l) - kT + (p_g v_l/N_l)$$

on using (19) This is equivalent to

$$p_g = (N_l kT/v_l\chi)\exp\{[\Phi_0/N_l kT] - 1\}, \tag{20}$$

since the term $(p_g v_l/N_l)$ may be neglected in comparison with kT. We note that this equation has the same form as (14). If we replace $v_l\chi$ by the "free volume" v_f, we may write equation (20) in the form

$$v_f^* = v_g^*\exp\{[\Phi_0/N_l kT] - 1\}. \tag{20 A}$$

This simple formula gives the "free volume" of the liquid per atom in terms of the specific volume of the gas

Equations (19) and (20) together determine the specific volume of the liquid as a function of temperature, but for moderate vapour pressures it is a good approximation to put p equal to zero in equation (19) as it is small compared with the other quantities involved, so that this equation alone gives the specific volume of the liquid as a function of temperature.

We may also determine the change of entropy and the heat of vaporization without difficulty. For the entropy of a substance is given by the well-known equation

$$TS = U - A,$$

so that from equations (12) and (17) the entropy per molecule of the liquid is given by

$$S_l^*/k = \frac{T}{\chi}\frac{\partial\chi}{\partial T} + \log\chi + \log(v_l/N_l) + \tfrac{3}{2}\log(2\pi mkT/h^2) + \tfrac{5}{2}, \tag{21}$$

---

* It is to be noted that Φ_0 and χ depend on N_l through (v_0/v_l) or (v_0^*/v_l^*).

while the entropy of the vapour, supposed to be a perfect gas, is given by

$$S_g^*/k = \log(v_g/N_g) + \tfrac{3}{2}\log(2\pi m kT/h^2) + \tfrac{5}{2}, \tag{22}$$

so that

$$(S_g^* - S_l^*)/k = \log(v_g^*/v_l^* \chi) - \frac{T}{\chi}\frac{\partial \chi}{\partial T}. \tag{23}$$

Now from equation (12) we see that $(kT^2/\chi)(\partial\chi/\partial T)$ is the change in potential energy of an atom due to its motion about its equilibrium position. We know that it will be equal to $\frac{3}{2}kT$ if the atom is moving like a harmonic oscillator This will not be far from the truth at high liquid densities and in any case the second term in equation (23) is always small compared with the first so that we may write approximately

$$\begin{aligned}(S_g^* - S_l^*)/k &= \log(v_g^*/v_l^* \chi) - \tfrac{3}{2}\\ &= \log(v_g^*/v_f^*) - \tfrac{3}{2}, \end{aligned} \tag{24}$$

The heat of vaporization is then given by

$$\begin{aligned}\Delta H^* = L^* = T(S_g^* - S_l^*) &= kT\log(v_g^*/v_l^* \chi) - \frac{kT^2}{\chi}\frac{\partial\chi}{\partial T}\\ &= kT\{\log(v_g^*/v_l^* \chi) - \tfrac{3}{2}\}. \end{aligned} \tag{25}$$

To the same approximation we can write equation (20A) in the alternative forms

$$\begin{aligned} v_f^* &= v_g^* \exp\left\{-\frac{\Delta H^*}{kT} - \tfrac{3}{2}\right\},\\ &= v_g^* \exp\left\{-\frac{\Delta V^*}{kT} - \tfrac{5}{2}\right\} \end{aligned}$$

3 DISCUSSION OF RESULTS

Equation (20) may be written in the form

$$p = \frac{N_l \Lambda}{v_0}\frac{kT}{\Lambda}\frac{v_0}{v_l}\frac{\exp\{(\Phi_0/N_l kT) - 1\}}{\chi}, \tag{26}$$

and hence, since $\Phi_0/N_l kT$ and χ both depend only on Λ/kT and v_l/v_0, and these quantities are given in terms of each other by equation (19) (with p equal to 0), we have

$$p = (N_l \Lambda/v_0)\Omega(\Lambda/kT), \tag{27}$$

where Ω is a dimensionless quantity, which is a function of Λ/kT only. It should be the same for all gases to which our equations are applicable. A few typical values of (v_l/v_0) and $\Omega(\Lambda/kT)$ for a range of values of Λ/kT, are given in Table I. As a guide to the size of the quantities involved it may be mentioned that Λ/kT is about 9 at the critical point and about 16 at the boiling-point for the simpler gases, and that for the inert gases $(N_l\Lambda/v_0)$ varies between 2×10^9 and 5×10^9 dynes/cm.^{2}. It is found that the calculated values of $\log\Omega$ are proportional to (Λ/kT) to within the errors of calculation, so that we can put the equation for the vapour pressure in the form

$$\log_e p = \log_e(N_l\Lambda/v_0) + 1{\cdot}916 - 0{\cdot}678(\Lambda/kT), \tag{28}$$

TABLE I

(Λ/kT)	(v_l/v_0)	$\Omega(\Lambda/kT)$	$\log_e \Omega$
12·8	1·118	0·00116	− 6·762
15·8	1·054	0·000159	− 8·749
18·0	1·026	0·0000352	−10·255
21·0	1·000	0·00000448	−12·317

for values of Λ/kT within the range 12·8–21·0. This is in agreement with the empirical fact that the vapour pressures of many gases can be expressed by an equation of this form. If we insert the values for Λ and v_0 calculated from the interatomic forces we find that the calculated vapour pressures expressed in dynes/cm.^{2} are expressed by an equation of the form

$$p = 10^6 \exp[A - B/T], \tag{29}$$

where the appropriate values of A and B are given in Table II. From these equations the boiling-points can at once be calculated and the values obtained are given in Table II. The theory can hardly be expected to apply to a light gas like hydrogen for reasons explained in Paper I, but it may be added that the boiling point comes out to be 26·4 instead of the observed value of 20·3.

TABLE II.* CALCULATED BOILING POINTS OF THE INERT GASES

			Boiling-point		$\dfrac{\text{Boiling-point}}{\text{Critical temp.}}$		Entropy of vaporization	
	A	B	Calc	Obs.	Calc.	Obs.	Calc.	Obs.
Neon	9·863	291·8	29·6	27·2	0·62	0·61	19·6	15·2
Argon	10·407	979	94·1	87·4	0·59	0·58	20·7	17·2
Nitrogen	9·950	786·3	79·0	77·2	0·61	0·61	19·8	17·3

[*] The boiling-points given here differ slightly from those quoted at the end of a previous paper (Lennard-Jones 1937), being the result of a revised calculation.

In this table we also give the entropy of vaporization in cal /deg This is given by using the well-known thermodynamic equation

$$\begin{aligned} \Delta S &= (v_g^* - v_l^*)(dp/dT) \\ &= (v_g^*)\,dp/dT \\ &= kTd(\log_e p)/dT, \end{aligned} \tag{30}$$

where p is the vapour pressure and then using the approximate equations (28) or (29) In fact from (28) we have

$$\Delta S = 0{\cdot}678k(\Lambda/kT), \tag{31}$$

and from the same equation we see that since $\log_e(N\Lambda/v_0)$ does not vary greatly from liquid to liquid, Λ/kT is practically a constant at the boiling-point Hence the entropy of vaporization should be approximately constant for all liquids to which this theory is applicable This is the familiar Trouton rule (probably more accurately referred to as the Pictet-Trouton rule) (Pictet 1876, Trouton, 1884) and it is satisfactory to find that the value given by the theory for the constant is practically the same as that used in the empirical rule Observation shows that the quantity has nearly the same value for a large number of gases.

The inner meaning of Trouton's rule becomes clear from an examination of equation (25). This equation shows that the latent heat of vaporization divided by the boiling temperature is determined largely by the ratio of the specific volume in the gas phase (v_g^*) at atmospheric pressure to the specific free volume in the liquid (v_f^*) at the same temperature and pressure The former of these at room temperature is of the order of $4{\cdot}10^{-20}$ c.c., while the latter may be taken to be of the order of 10^{-24} c c. We see at once that $\log_e(v_g^*/v_f^*)$ is of the order of 10 and so L/T is of the order of 20, as given by observation.

Trouton's rule has been modified by Hildebrand to the statement that the entropy of vaporization is the same for all liquids at temperatures at which the concentration in the gas phase has a given value. The significance of this becomes clear from equation (28) This equation can be written in the form

$$\log_e(v_o^*/v_g^*) = 11{\cdot}916 - (0{\cdot}678)(\Lambda/kT) + \log_e(\Lambda/kT). \tag{32}$$

Now v_o^* is the specific volume of the liquid when the molecules are at the same distance apart as in the diatomic molecule This will not vary by more than a factor of 3 or 4, so that when v_g^* is given, that is the concentration in the vapour phase given, Λ/kT will be approximately the same for all liquids, and hence the entropies of vaporization will be the same.

Hildebrand's rule would according to this theory be correct, if all substances had the same number of molecules per unit volume at absolute zero.

The actual values of Λ/kT at the boiling-point for the liquids in Table II lie between 16 and 17 so that remembering that $\Lambda = 12\phi_0$, we have

$$kT_b \doteq 0{\cdot}06\Lambda = 0{\cdot}72\phi_0,$$

where T_b is the temperature at the boiling-point and ϕ_0 was defined to be the absolute value of the mutual potential energy of two molecules of the substance in their equilibrium position.

In the table we also give the ratio of the calculated boiling-points to the critical temperatures calculated in Paper I, and the observed values of these ratios.

Finally we may note that if we had used the same partition function as in Paper I, that is, without the correcting factor e^N, the calculated vapour pressures would have been higher by a factor e, and the boiling-points about 9 % lower They would then have been rather below the experimental values instead of above them.

We are indebted to the Department of Scientific and Industrial Research for a grant to enable this work to be carried out.

SUMMARY

This paper extends to liquids the theory which was given recently by the authors for the equation of state of a gas at high densities. A direct calculation is made of boiling points in terms of interatomic forces and the numerical results for the inert gases are in satisfactory agreement with the observed values A theoretical interpretation is given of Trouton's empirical rule connecting the heat of vaporization with the boiling temperature and also of Hildebrand's modified form of it Calculations are made of the vapour pressures and heats of vaporization of the inert gases.

REFERENCES

Eyring 1936 *J. Chem Phys* **4**, 283.
Eyring and Hirschfelder 1937 *J Phys. Chem.* **41**, 249.
Fowler 1936 "Statistical Mechanics" Camb Univ Press
Hirschfelder, Stevenson and Eyring 1937 *J Chem. Phys* **5**, 896.
Kincaid and Eyring 1937 *J Chem. Phys.* **5**, 587
Lennard-Jones 1937 *Physica*, Nov

Lennard-Jones and Ingham 1925 *Proc. Roy. Soc.* A, **107**, 636
Lennard-Jones and Devonshire 1937 *Proc. Roy. Soc.* A, **163**, 53.
Mayer 1937 *J. Chem Phys.* **5**, 68
Mayer and Ackermann 1937 *J. Chem. Phys.* **5**, 74
Newton and Eyring 1937 *Trans. Faraday Soc* **33**, 73
Pictet 1876 *Ann. Chem Phys* **183**, 180.
Trouton 1884 *Phil. Mag* **18**, 54.
Wheeler 1934 *Indian J. Phys.* **8**, 521
— 1935–6 *Proc Ind Acad Sci* **1**, 105, 795; **2**, 1, 466, **4**, 291, 298

The nature of the penetrating component of cosmic rays

By P. M. S Blackett, F R S

(*Received* 18 *December* 1937)

1 Introduction and summary of conclusions

The measurements by Neddermeyer and Anderson (1937) of the absorption of cosmic-ray particles of low energy by metal plates differ in certain respects from those by Blackett and Wilson (1937). The former results showed that, in the energy range $1 \cdot 2 \times 10^8$ to 5×10^8 e-volts, two types of particles exist, an absorbable group assumed to behave as theory predicts of electrons and a much more penetrating group, attributed provisionally to heavier particles.

On the other hand, we found that all the rays with energy under 2×10^8 e-volts were absorbed like electrons, while for rays of greater energy the average energy loss was very much less. Though a very few energetic particles were found to have a high energy loss, insufficient evidence was then available to justify classifying them as of a nature distinct from the less absorbable rays. Thus we obtained definite experimental evidence that the energy loss of the great majority of the rays varies rapidly with their energy. We concluded, therefore, that the energy loss of a normal electron varies with its energy. We now believe this to be probably false, since the success of the cascade theory of showers, in explaining the transition curve in the atmosphere, and a large part, at any rate, of the phenomena of the transition curves of showers and bursts, has provided fairly strong evidence that there must be a very few energetic rays at sea-level, which have the full radiation loss of electrons, even in heavy elements It follows that the great majority of the rays, for which the energy loss certainly varies rapidly with energy, are probably not normal electrons. We therefore agree with the view of Neddermeyer and Anderson that it is likely that there are two types of particles present, though the difference in behaviour only exists for energies over 2×10^8 e-volts.

The experimental work described in this paper was carried out mainly to investigate the discrepancy mentioned above between the two sets of experimental results. It will be shown that our former conclusion, that nearly all the rays with energy under 2×10^8 e-volts are absorbed like radiating electrons, has been fully confirmed. Further, the assumption of the previous paper that the penetrating rays actually become absorbable like normal electrons, when their energy falls much below about 2×10^8 e-volts, is also shown to be correct

These further results, together with the implications of the cascade theory, lead therefore to the conclusion that the cosmic-ray beam at sea-level consists of a few fully radiating electrons, together with a large number of particles, which are very penetrating when energetic, but which apparently become indistinguishable from radiating electrons when their energy falls much below 2×10^8 e-volts.

2 THE MEAN ENERGY LOSS AS A FUNCTION OF ENERGY

Measurements have been made of the energy loss of cosmic rays in the following plates

	t	$l = t/\lambda_0$
Lead	0 33 cm.	0 82
Lead	1 0 cm	2 50
Gold	2 0 cm.	8·5

The thicknesses (t) are given also in terms of the fundamental units λ_0 of the cascade theory, which are 0·40 *cm* Pb and 0·24 *cm* Au (Bhabha and Heitler 1937). The magnetic field used was either 3330 or 10,000 gauss.

In order to compare the results with the theory of the radiation loss, it is convenient to subtract from the measured energy loss $(E_1 - E_2)$, the energy E_i estimated to be lost in ionization and excitation. From Table II of the paper by Bethe and Heitler (1934), this can be estimated as approximately 15×10^6 e-volts per cm. Pb, for energies from fifty to a few hundred million volts. Thus the value of the relative energy loss is calculated from the expression

$$R = (E_1 - E_2 - E_i t)/\bar{E}t, \tag{1}$$

where t cm. is the thickness of the plate, and $\bar{E} = \frac{1}{2}(E_1 + E_2)$ is the mean of the measured energy E_1 above, and the energy E_2 below the plate *

The results of the measurements for the two lead plates are given in Table I. The results with the gold plate are given in § 3

TABLE I ENERGY LOSS IN LEAD PLATES

Absorber	Range of E 10^6 e-volts	R from equation (1) No	R from equation (1) (E_1)	R from equation (1) No	R from equation (1) $(\bar{E})$	R_c from equation (2) (E_1)	R_c from equation (2) $(\bar{E})$
0.33 cm lead, 3300 gauss	0–100	29	2.01 ±0.33	20	2.42 ±0.37	2.84 ±0.47	3.09 ±0.48
	100–200	13	1.00 ±0.26	9	1.05 ±0.31	1.13 ±0.29	1.18 ±0.34
	200–300	10	0.98 ±0.28	9	0.46 ±0.18	1.06 ±0.28	0.53 ±0.18
	300–400	13	0.024 ±0.075	11	0.16 ±0.08	0.069 ±0.075	0.21 ±0.08
	400–500	6	0.19 ±0.40	6	0.10 ±0.21	0.22 ±0.40	0.14 ±0.21
	500–700	12	0.14 ±0.16	13	0.09 ±0.14	0.16 ±0.16	0.12 ±0.14
1.0 cm lead							
Mainly 3300 gauss	0–100	19	0.99 ±0.08	20	1.04 ±0.09	2.01 ±0.16	2.11 ±0.18
	100–200	5	1.27 ±0.25	(2)	—	1.80 ±0.35	—
Mainly 10,000 gauss	200–500	10	0.12 ±0.054	10	0.44 ±0.14	0.16 ±0.072	0.50 ±0.16
	500–700	13	0.25 ±0.075	10	0.039 ±0.046	0.26 ±0.08	0.067 ±0.046

Of the tracks with the thin plate, twenty-eight are from the previous paper by Blackett and Wilson (1937) and the remaining fifty-five are new. Of those with the thick plate, nearly all with energy under 2×10^8 e-volts are new, and most of the others are from the previous paper. All the earlier results were obtained with counters above and below the chamber, thus giving a bias against large losses. For nearly all the new results, all the counters were above the chamber. This is the better arrangement and is essentia for the low energy rays when using a thick plate. The actual results obtained in both systems were essentially the same. For most of the photographs

---

* The error of a single energy measurement is given by $\delta E = E^2/E_m$ (Blackett 1937*a*), where E_m is about 4×10^9 e-volts for the tracks in a field of 3300 gauss and about 1.2×10^{10} e-volts for 10,000 gauss. The error δR of R is approximately $\sqrt{2E/E_m t}$ For instance, for $E = 5 \times 10^8$ e-volts, we have $\delta R = 0.4$ for the thin lead plate, and 0.14 for the thick lead plate For low energies, however, the errors are rather larger than those given by the above expression The corresponding value of E_m in the measurements of Neddermeyer and Anderson (1937) appears to have been about 6×10^8 e-volts, with a field of about 7000 gauss.

For $E > 3 \times 10^8$ e-volts the error of a single energy-loss determination is comparable with the real mean energy loss, but for $E < 2 \times 10^8$ e-volts, the error is much less than the real mean loss, and the error of R is then mainly due to the statistical variation of the number of rays observed with given energy loss a lead screen about 1·5 cm. thick was placed over the top counter to increase the number of recorded showers.

The two values of $\bar{R}$ listed for each energy range are obtained by grouping the particles differently. In the column headed (E_i), the value of R is the mean of the values of R for all tracks whose initial energy E_i lies within the stated limits, while the column headed ($\bar{E}$) refers to all tracks which have their mean energy $\bar{E}$ in the stated limits. The latter method was used in the previous paper as it is more reliable, since the mean error of $\bar{E}$ is $1/\sqrt{2}$ times that of E_i. But the classification by initial energy E_i shows up better the sharp fall of $\bar{R}$ with energy.

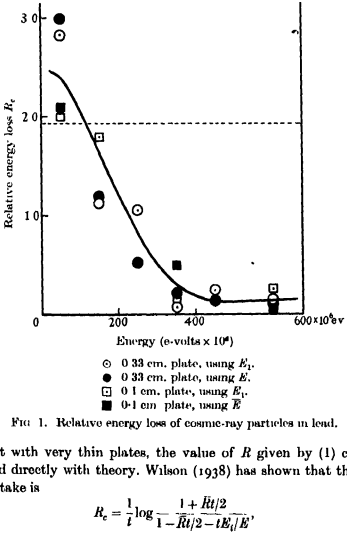


FIG. 1. Relative energy loss of cosmic-ray particles in lead.

Except with very thin plates, the value of R given by (1) cannot be compared directly with theory. Wilson (1938) has shown that the correct value to take is

$$R_c = \frac{1}{t}\log\frac{1+\bar{R}t/2}{1-\bar{R}t/2-tE_i/\bar{E}}, \qquad (2)$$

where E_i is the energy lost in ionization These corrected values are given in the last two columns, and are shown graphically in fig 1. The open circles, etc., give the data obtained by the classification by E_1, and the black circles, etc., that by $\bar{E}$ All the open circles, etc , represent one way of analysing all the tracks, and the black circles, etc., another way of analysing the *same* data These results are in close agreement with those previously reported. The sharp fall of R for $E \simeq 2 \times 10^8$ e-volts is fully confirmed. For lower energies the observed values are comparable with the expected loss of an electron (dotted curve), but for greater energies are much less than this. For the lowest energies, the observed values are greater than the theoretical loss, in agreement with the results of Turin and Crane (1937) for electrons of about 10^7 e-volts These high values are possibly due to scattering The true curve must be steeper than that observed owing to the effects of errors, but how much steeper it is difficult to say without further measurements.

3. The distribution of energy losses for $E < 2 \times 10^8$ e-volts

Since for this energy region the mean energy loss is comparable with the theoretical value for electrons, it is important to compare also the distribution of energy losses with the theory. Wilson (1938) has derived from the expressions given by Heitler (1937) the following results If t be the thickness of the absorber, and λ_0 be the fundamental unit of length of the cascade theory, and if $\alpha = Rt/2$, where R is the relative energy loss of a track as defined by (1), then the chance of an energy loss between α and $\alpha + d\alpha$ is

$$w(\alpha)\, d\alpha \propto \frac{d\alpha}{(1+\alpha)^2}\left[\log\left\{\frac{1+\alpha}{1-\alpha}\right\}\right]^{\frac{t}{\lambda_0}-1}. \qquad (3)$$

The curves in figs 2 and 3 show this expected distribution of energy loss for the 0·33 and the 1·0 cm. lead plates ($l = t/\lambda_0 = 0{\cdot}82$ and $l = 2{\cdot}50$)

The points on the figures represent the observed distribution of energy loss of all the rays with energy less than 2×10^8 e-volts. The agreement of the observation with theory is quite as good as can be expected, in view of the approximation of the theory and the statistical and other errors of the experiments With the thin plate, fig. 2 shows that the chance of a large loss of energy is much less than the chance of a small loss, and the distribution curve is of such a shape as to make easy the detection of particles with energy loss greater than that of electrons, if such exist Fig 2 shows no evidence of such particles. The explanation of the high value for $R \simeq 4{\cdot}5$ is given later

On the other hand, with the thick plate, fig 3 shows that the chance of a small loss is much smaller than that of large loss, and so the conditions are favourable for the detection of rays with an energy loss that is considerably less than that of electrons. No sign of an appreciable number of such particles is found.

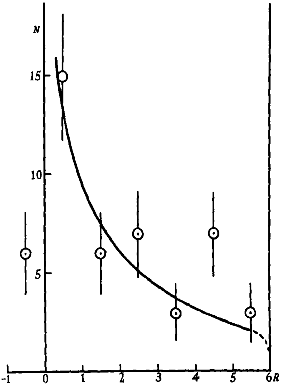


Fig 2 Distribution of energy loss in 0·33 cm lead plate $H = 3300$ gauss $E < 200$ e-volts × 10^6.

When a 2 cm. gold plate is used, it is not possible to measure the energy loss of radiating electrons, since they are almost all totally absorbed. But such a plate is useful to separate out radiating electrons from more penetrating particles A full analysis of the photographs taken with the gold plate is not yet finished, but the following are some preliminary results. Table II gives a list of the energies and sign of all rays observed with energies between 5×10^7 and 5×10^8 e-volts and shows in a very striking way the change of energy loss for $E \cong 2 \times 10^8$ e-volts

Of the thirty-five tracks, nineteen stop in the plate, and all are under $2{\cdot}5 \times 10^8$ e-volts. Of the sixteen which traverse the plate, all but one are

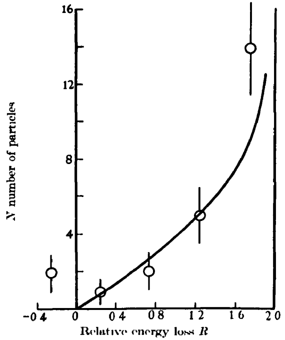


FIG 3 Distribution of energy loss in 1·0 cm lead plate. $E < 200$ e-volts × 10^6

TABLE II. ABSORPTION BY A 2 CM GOLD PLATE

Initial energies in million electron volts	
Rays which stop in plate	Rays which traverse plate
+ 53	+ 165
+ 62	+ 227
− 65	− 232
− 74	+ 258
− 78*	+ 261*
− 83	+ 275*
− 88	− 320
+ 95	− 322
− 96	− 330*
+ 121	+ 330
− 132*	− 430*
+ 140	+ 460
+ 165	+ 480*
− 175	+ 470*
− 186	+ 480
− 188	+ 490*
+ 209*	
− 224*	
− 242	

over $2{\cdot}2 \times 10^8$ e-volts, and lose only a small fraction of their energy. The one which has a lower energy ($1{\cdot}65 \times 10^8$ e-volts) actually loses about half its energy, and can be interpreted in two different ways, according as whether the curve of fig 1 is considered to represent the actual variation of energy loss of each particle, or as representing the result of a sudden change of radiating power occurring at different energies for different particles (see § 8) In the first case the track represents a less than average energy loss of a particle of mean energy loss considerably less than that of an electron, in the second, a particle which makes the sudden transition abnormally late.

The tracks marked with an asterisk were made with a 5 or a 15 cm. lead absorber between the counters, which were all above the chamber (see § 7) For the rest, there was no heavy absorber over the chamber

Figs. 4 and 5 show graphically the measured relative energy loss R of each individual track, plotted against its initial energy E_1, for the two lead plates * The data from the previous paper are used to extend fig. 5 up to $E = 2 \times 10^9$ e-volts Each point which corresponds to a ray which entered the chamber accompanied by one or more other rays is considered to be part of a shower and is marked by two vertical lines It is seen that for $E < 2 \times 10^8$ e-volts, the distribution of shower particles is sensibly the same as that of the single particles, and further, the number of positive particles is sensibly the same as the number of negatives We conclude from these results, and from those of § 2, that rays with $E < 2 \times 10^8$ e-volts are all positive and negative electrons, with approximately the full radiation loss predicted by theory. If any more (or less) penetrating particles are present in this energy range, their number, in our experiments, cannot exceed a few per cent

This conclusion is in conflict with that of Neddermeyer and Anderson (1937), who find a strong penetrating group extending down to nearly $1{\cdot}2 \times 10^8$ e-volts. The possibility that the discrepancy could be due to different amounts and nature of material over the chamber seems excluded, since we find no sign of the penetrating group of low energy, either with no

---

* The fact that, for the 0·33 and the 1·0 cm plates, the maximum values of R (corresponding to complete stoppage) are less at low energies than 6 and 2 respectively, is due to the subtraction of the ionization loss ($E_i t$) from the observed energy loss (equation (1)). For tracks that stop, and so do not traverse the full distance t, this subtracted energy is clearly too great by some amount that, for any one track, cannot be determined. The high value of one point in fig 2 is to be attributed partly to this cause, and partly to the systematic error introduced by the scattering in the plate of the slow rays, and by their deflexion from the vertical by the magnetic field. Both effects increase the real path in the absorber, and so give too many high values of R. Owing to the subtraction of the ionization loss, the high numbers occur for $R \cong 4{\cdot}5$ in fig. 4, instead of for $R \cong 6$

heavy material, with 1·5 cm Pb, with 15 cm Pb, or with 8 cm. Cu over the chamber It will therefore be taken as proved that there is no large penetrating group with $E < 2 \times 10^8$ e-volts.

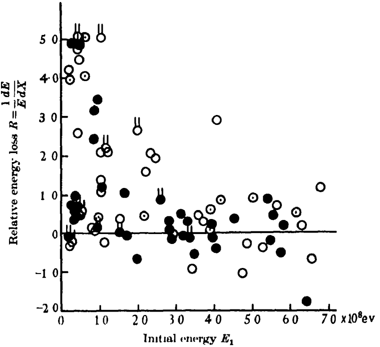


FIG 4. Distribution of relative energy loss in 0 33 cm lead plate

⊙ Single tracks
⦶ Tracks associated with other particles
● Negative particles.
○ Positive particles.

4. THE RAYS WITH ENERGY GREATER THAN 3×10^8 E-VOLTS

From figs 1 and 5 it is seen that the mean energy loss of the rays with $E > 3 \times 10^8$ e-volts, is much less than that of radiating electrons, and yet they cannot be mainly protons.* However, a few particles do lose a large fraction of their energy. For instance, in fig 5 six particles out of fifty, in the range 3×10^8 to 2×10^9 e-volts, have values of R greater than 0·6, while in fig. 4 one particle out of thirty-five in the range 3×10^8 to 7×10^8 e-volts has a value of R greater than 2·0. Thus rather less than 10 % of the particles with energy between these limits are heavily absorbed as would be expected of electrons

---

* Similar evidence against this possibility has been obtained by Crussard and Leprince Ringuet (1937) and by Street and Stevenson (1937) The point marked P in fig. 5 is the only recognizable proton among the tracks described here.

Between the energy of 2×10^8 e-volts, below which all the rays are absorbed approximately like electrons, and the energy of 3×10^8 e-volts, above which only a small fraction are absorbed electronically, lies a transition region which needs further investigation, since the results for the two plates differ somewhat.

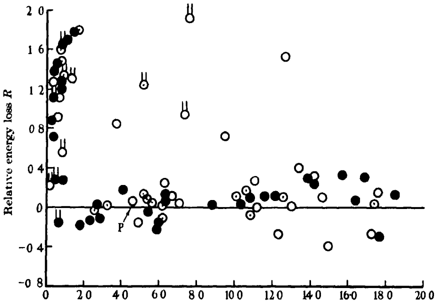


FIG. 5. Distribution of relative energy loss in a 1 cm. lead plate.

● Negative particles.

○ Positive particles

Of the seven absorbable, but energetic, particles shown in figs 4 and 5, three are in showers starting in the lead over the counters, while all but one of the penetrating group are single. Street and Stevenson (1936) were the first to recognize clearly the great difference between the behaviour of single particles and those in showers. As their photographs were taken with no magnetic field, it was not possible to be sure that this difference in behaviour might not be due to a difference of energy rather than of kind. As has been mentioned, Neddermeyer and Anderson (1937) found evidence for the existence of both an absorbable and a penetrating group in the energy range $1{\cdot}2 \times 10^8$ to 5×10^8 e-volts and concluded that the two groups correspond to normal and to "heavy" electrons. Our results confirm the probable existence of the two groups, but only in the energy region above about 2.5×10^8 e-volts. The scanty data in fig. 5 seem to represent still the only direct evidence available at present that the two groups exist in the same energy range, and even here the accuracy of the measurements is not as high as would be desirable. The distribution of the six absorbable rays is about what is expected for electrons. The fact that all are positive may be due to chance, or to a possible asymmetry between positives and negatives among the shower particles. Such an asymmetry, if established, would prove a grave difficulty for the cascade theory.

The strongest evidence for the real existence of the two groups in the same energy range is obtained indirectly from the cascade theory of showers, as developed by Bhabha and Heitler (1937) and by Carlson and Oppenheimer (1937). The explanation of the atmospheric transition curve given by Heitler (1937) leaves little doubt that there are some electrons that radiate fully in light elements up to 10^{10} volts or more. For heavy elements the position is less clear, but the evidence from the shower transition curves suggests that some electrons radiate fully also in heavy elements up to high energies. Further evidence for this is given in § 9.2. The assumption that the radiation formula does not break down clearly necessitates the assumption of two groups in the same energy range, in order to account for the penetrating group, since this certainly exists over a very wide range of energy.

5. THE BEHAVIOUR OF THE PENETRATING RAYS

Apart from the six rays showing a large absorption, the rays in fig. 5 with $E > 3 \times 10^8$ e-volts are about equally distributed between positives and negatives, and their mean energy loss is perhaps of the order of 1/16 that of electrons.

Two questions immediately present themselves. What happens to the rays of this penetrating group, when their energy falls below 2×10^8 e-volts, and where do the electrons and positrons with $E < 2 \times 10^8$ e-volts come from? The most obvious assumption, which was made in the previous paper by Blackett and Wilson (1937), is that the penetrating rays of high energy become, as they lose energy, the absorbable rays of low energy, that is that their energy loss varies with their energy, and approaches that of electrons at low energies.

Though strong evidence that this view is correct will be given in §§ 7 and 8 it is interesting to consider what is involved in the attempt to explain the facts by means only of particles with a constant relative energy loss, for instance by "heavy" electrons of constant rest mass. One would then have to assume that such particles stopped suddenly as their energy fell below 2×10^8 e-volts. No sign of such a process has been observed. In fact, it is shown in § 3 that there is no sign in this region of any process of absorption of *large* fractions of the energy of a particle, other than that characteristic of radiating electrons. If such a catastrophic process took place it could hardly fail to have been observed. For instance, in fig. 4, one would expect to find some particles of energy about 2×10^8 e-volts, which stop completely, but there are none.

6 THE EVIDENCE FROM THE LOW ENERGY END OF THE SPECTRUM

No satisfactory investigation of this end of the spectrum, or how it varies below different absorbers, is available. However, Table III gives some rather scanty evidence from photographs taken, (*a*) under no heavy absorber, (*b*) under different lead absorbers, varying from 3 mm. to 5 cm. in thickness. The numbers are not very reliable, owing to the effect of the magnetic field in deflecting away the slow rays. The number of these is therefore certainly underestimated. For the fraction of the spectrum over 10^9 e-volts, the data given by Blackett (1937*a*) have been used.

TABLE III. LOW ENERGY END OF SPECTRUM

Energy range ($\times 10^8$ e-volts)	Number of tracks		Total	Fraction
	No absorber	Lead absorber		
0–1	18	30	48	0·05
1–2	27	18	45	0·05
2–3	28	11	39	0·04
3–4	32	21	53	0·06
4–5	16	25	41	0·04
5–6	18	22	40	0·04
6–7	19	15	34	0·04
7–8	23	7	30	0·03
8–9	21	15	36	0·03
9–10	21	15	36	0·04
10–∞	—	—	(560)	(0·58)
			(960)	(0·42)

For the case of no heavy absorber, the spectrum is fairly flat, but there is a sign of a minimum for $E \simeq 2\times 10^8$ e-volts under the lead absorber. This minimum, which was mentioned in the previous paper, is just what would be expected, owing to the rather sudden increase of the energy loss with decreasing energy, as shown in fig. 1. The minimum would not be expected with no heavy absorber, even if, as is possible (see Wilson 1938), the radiation loss in air also increases suddenly at about the same energy, because of the much greater value of the ionization loss relative to the radiation loss in air in comparison with lead. In air, therefore, one would expect, either no minimum at all, or only a small one. The expected relation between a minimum in the spectrum and a sudden increase of energy loss has already been discussed by Blackett (1937*a*) in another connexion.

This sudden reduction of the intensity of the energy spectrum under a heavy absorber is also very well shown by the results for the tracks emerging from the gold plate. For if those tracks listed in the right-hand column of Table II are considered, it is seen that the number per unit energy range under 2×10^8 e-volts is much less than the number with higher energy. Since this change of intensity *under* a heavy absorber is just what would be expected if the penetrating particles actually become absorbable, it must be considered as providing strong evidence that the change of property actually does take place.

The observed fraction of the rays with energy greater than 3×10^8 e-volts, which are radiating electrons, depends on the number of showers observed, and this of course depends on the nature of the absorber over the apparatus (transition curve). From the data of figs. 4 and 5 this fraction in our experiments for the range 3×10^8 to 2×10^9 e-volts is about 10 %. For the whole range from 3×10^8 e-volts to infinity the fraction must be less than this, say 5 % or less, owing to the difference in the shape of the energy spectra (see § 9·1). These figures refer to the condition under a lead absorber of about 1·5 cm. thickness. In air, that is with no absorber, the fraction must be less than 1·5 % (see § 9·1). On the basis of these results we can construct fig. 6 to represent, highly schematically, the low energy end of the spectrum. To simplify the discussion, the energy loss has been assumed to change suddenly at the critical energy.

The approximate numbers of rays in the three main parts are as follows:

	Particles	Energy range	Fraction	Absorption group
A	Electrons Positrons	$E < 2{\cdot}5 \times 10^8$ e-volt	about 20 %	Soft
B	Ditto	$E > 2{\cdot}5 \times 10^8$ e-volts	in air 1·5 % under Pb 5 %	
C	Unknown	$E > 2{\cdot}5 \times 10^8$ e-volts	about 80 %	Hard

Now if the hard group C does disappear at about 2×10^8 e-volts, the soft groups A and B must, independently of C, form the spectrum of the radiating electrons. But such a spectrum, shown by the thick line, is quite inconsistent with the cascade theory (Heitler 1937). This theory shows that for $E > E_0$, where $E_0 \gtrsim 1\cdot3 \times 10^8$ e-volts for air, the number of rays between E and $E + dE$ falls off about as $E^{-2\cdot5}$, but for $E < E_0$ is nearly constant, that is more, as is shown by the horizontal dotted curve. We conclude therefore that the radiation theory can only explain the part B′ of group A, but not the rest C′, which must therefore be recruited from group C. It follows that the rays of the penetrating group, whatever their nature when energetic, are apparently absorbed like electrons when their energy falls much below 2×10^8 e-volts.

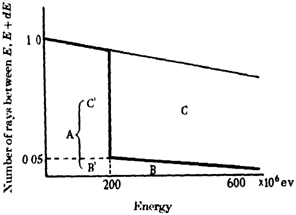


FIG. 6. Schematic representation of low energy end of spectrum.

7. DIRECT EVIDENCE FOR THE CHANGE OF RADIATION LOSS OF PENETRATING PARTICLE

In order to test directly the conclusion of the previous section that a penetrating particle apparently becomes an electron when it slows down, a series of photographs was taken with three counters over the chamber, and between them a lead absorber, either of 5 or 15 cm. thickness. This absorber must cut out completely the absorbable rays (except for a few of high energy, which will appear as showers), letting only the penetrating rays reach the chamber. Thus if any ray enters the chamber singly with energy less than about 2×10^8 e-volts and is absorbed by the 2 cm. gold plate in the normal manner for electrons, it is proved that a penetrating ray has become, or has given rise to, an electron. The single tracks taken with this arrangement are included in Table II, showing the results with the gold plate, and are those marked with an asterisk. It will be seen that there are four rays, which have traversed the lead plate, but which stop in the gold plate. One would only expect a few of such rays, since the chance that a penetrating particle will emerge from the lead plate just in the right energy region is quite small. These results, though few in number and therefore only of a preliminary character, give therefore some direct evidence that a penetrating ray actually becomes absorbable like an electron as it slows down.

Since the number of normal radiating electrons observed in our experiments is only a small fraction of all the rays, it follows that the experimental curve (fig 1) giving the dependence on energy of the energy loss of all the rays, actually represents fairly accurately the energy loss curve of the penetrating rays alone. Whether this change of their property of radiating energy happens gradually or suddenly can only be settled by further experiments On the whole, the existing evidence suggests a fairly rapid, but not discontinuous, rise of radiation loss as E falls below 3×10^8 e-volts It is possible that the full radiation loss of an electron is not obtained till its energy is less than 10^8 e-volts

Each penetrating ray, when it reaches an energy of less than 2×10^8 e-volts will produce a small shower, since it then behaves like a radiating electron This may be the origin of the smaller showers, observed by counter experiments to be due to the hard component.

8 DISCUSSION OF THE POSSIBLE NATURE OF THE PENETRATING RAYS

Clearly the first demand, which any hypothesis as to the nature of the penetrating component must satisfy, is that it must explain the observed rapid change in the behaviour of the particles when their energy is about $2\cdot5\times10^8$ e-volts. It will be essential to discover whether this change is sudden or gradual, and whether the energy at which it occurs varies with the absorber Recent results by Wilson (1938) appear to show a somewhat less rapid variation with energy in copper than in lead, but the variation with atomic number Z is not nearly as large as was suggested previously *

---

* In a previous paper Blackett and Wilson (1937) showed that the observed variation of the mean energy loss with energy was in rough agreement with a tentative theory of Nordheim (1936), to explain the then supposed breakdown of the radiation loss of all electrons I am indebted to Dr E J Williams for pointing out to me that an alternative theory put forward previously by him (1934) fits the results for lead equally well, and the new results for lighter elements much better For the former theory gives a large change of radiation property with Z ($E_c \propto Z^{-5/3}$), and the latter a much smaller variation ($E_c \propto Z^{1/3}$) The earlier results with aluminium, which supported the former expression, had a rather large probable error and were probably

It is clear that the penetrating rays cannot be heavy electrons of constant rest mass, or of any mixture of such heavy electrons, since the mean radiation loss varies rapidly. They might, however, consist of heavy electrons with a variable rest mass, for instance "excited" electrons which go to their normal state when their energy drops below the critical value. Alternatively, the heavy particle may in some way give rise to, rather than become, the light particles. It is obvious from relativity considerations that the change of property cannot occur in free space, but must occur during some type of collision.

To explain the rather rapid change of property with energy, one might postulate some new kind of resonance phenomenon between the particles and the nuclear fields of the absorbing atoms This resonance would then have to occur for a critical energy E_c, which probably increases slowly as the atomic number Z decreases.

A rather strong argument for the assumption that it is a difference of mass which distinguishes the two types of particles, lies in some results, which will be described elsewhere, which show that though the radiation from these particles is small, their scattering agrees with the theoretical value for very fast particles Much the simplest way of reconciling a subnormal radiation loss and a normal scattering is to assume a greater mass, since other possible ways of reducing the radiation loss are likely also to reduce the scattering.

This type of explanation will only be plausible if the change of property of each particle proves to be sudden If the gradual change shown in fig 1 really represents the behaviour of the individual particles, such an explanation would be untenable

From fig 1 it is seen that the radiation loss of the penetrating rays falls to about 1/16 of that for electrons, or perhaps less, for energies of about 4×10^8 e-volts. This gives about four times the electronic mass as the lower limit of the rest mass of the particles in this energy region. It seems improbable that they can be heavier than say, 100 times the electronic mass,

given undue weight (see Wilson 1938) It is however doubtful if Williams' theory is capable of explaining the very rapid fall of R now observed

[*Note added in proof* In a recent paper March (1937) has given an explanation of the observed drop in the energy loss, as indicating that there exists in nature a lower limit γ for the wavelength of a photon that can react with matter at rest. From the observed critical energy of about 2×10^8 e-volts, it follows that $\gamma \cong 2r_0$, where r_0 is the classical radius of the electron. This theory is therefore closely related to the suggestion of Williams (1934) that the breakdown of the radiation loss is connected with the classical electron radius But since the theory must apply to all electrons, it is inconsistent, unless further assumptions are made, with the facts explained so well by the theory of cascade showers.]

since momentum considerations would then prevent them giving rise to sufficiently fast electrons to explain the group C' of fig. 6

Whatever the mechanism of the change of radiative property, it must occur without abnormal scattering, since the observed scattering is, at any rate, approximately normal at all energies From considerations of momentum and energy transfer, no simple collision mechanism can easily be imagined which will transform a heavy particle into a light one without appreciable scattering

If, as is possible, the energy loss will be found to be a continuous function of the energy of the particle, one will be forced to some other type of explanation, perhaps to one in which the penetrating rays are supposed to be of electronic rest mass, but are distinguishable from ordinary electrons by some new physical property having the effects of making the radiation loss a function of energy Such a theory would then resemble closely the various breakdown theories discussed in the former paper, in that these theories were formulated to explain the then supposed variation of energy loss of electrons with E and Z It is now, however, clear that, if any rays obey such a "breakdown" theory, it is the penetrating rays alone (that is, on this view, the abnormal and not the normal electrons) Of course the supposed mechanism behind the variation of radiation property with energy would have to be altered, so as to depend on the unknown physical property distinguishing the penetrating rays from the normal electrons But some new mechanism, not contained in the present structure of quantum mechanism, is certainly necessary to explain all the facts.

9·1. THE FRACTION OF COSMIC RAYS AT SEA-LEVEL WHICH ARE RADIATING ELECTRONS

The fraction of the rays *in air* which are normal electrons is much less than is indicated by the spectrum of absorbable particles with 3×10^8 e-volts $< E < 2 \times 10^9$ e-volts in fig 5 For the photographs were taken under lead, and further all *counter systems are strongly selective for* showers. An independent estimate of the upper limit of the fraction of rays in air which are electrons, can be obtained from the observed number of bursts For, on the cascade theory, almost every energetic electron will produce a burst

For instance, from the data given by Ehrenberg (1936), it can be calculated that the ratio of the number of bursts, with more than ten rays, to the number of singles, is between 5×10^{-4} and 10^{-3}. Again, from the data given by Montgomery and Montgomery (1935*a*, fig 3) the ratio of bursts ($N > 10$)

under a 1 cm lead plate to singles, is found to be rather less than 10^{-3}. As a rough result we will therefore assume the ratio to be 10^{-3} for a lead plate 1·2 cm thick

Now from the data given by Heitler (1937), a shower of ten particles is produced in a lead plate of thickness 1·2 cm ($l = 3$) by an electron of energy parameter $y = 5{\cdot}2$ Since the critical energy for lead is 10^7 e-volts, this value of $y = \log(E/E_0)$ corresponds to an energy of $1{\cdot}8 \times 10^9$ e-volts Since all electrons of higher energies will produce larger showers (neglecting fluctuations), the fraction of the cosmic-ray beam which consists of electrons with energy greater than $1{\cdot}8 \times 10^9$ e-volts, must be equal to the number of bursts with $N > 10$, expressed as a fraction of the number of singles, this ratio is then about 10^{-3} To obtain the total number of electrons with $E > 3 \times 10^8$ e-volts, we can use the theoretical energy distribution given by Heitler (1937, p. 276). This is, that the number of electrons with energy greater than E is roughly proportional to $E^{-1{\cdot}5}$. Thus the number of electrons with $E > 3 \times 10^8$ e-volts is $10^{-3}(18/3)^{1{\cdot}5}$, that is about 1·5 % We conclude therefore that at sea-level in air the fraction of the rays which are normal electrons (or positrons), with energy greater than 3×10^8 e-volts, is not more than 1·5 %. This estimate is an upper limit, as some of the observed bursts are probably produced by the penetrating component

It is interesting to compare the energy spectrum in air of the absorbable and the penetrating rays As is mentioned in § 6, the spectrum (that is the numbered rays with energy between $E, E+dE$) of the former should be roughly constant for $E < 1{\cdot}3 \times 10^8$ e-volts and should then fall off about as $E^{-2{\cdot}5}$, or according to Carlson and Oppenheimer (1937) as E^{-2}

Now from Table III, and from the results given by Blackett (1937*a*), the spectrum of the penetrating rays is seen to be nearly constant for $E < 10^9$ e-volts and then falls off about as E^{-2} or rather faster Thus the spectrum of the penetrating rays is rather similar to that of the electrons, but the value of the energy at which the change of slope occurs is about seven times larger

Since the number of energetic electrons at sea-level is very small, it follows that the observed absorption at sea-level, except for small thicknesses of absorbers, is almost entirely due to the penetrating component Thus the results obtained by Blackett (1937*b*) for the energy loss of all the rays at sea-level, by comparing the observed absorption and energy spectra refer, effectively, only to the penetrating component, and are very little affected, except for small thicknesses of absorber, by the presence of the few energetic electrons. Since it is probable that the straggling of the penetrating component is small, the calculated energy loss, which is based on the assumption of a range-velocity relation, should be reliable. Thus the mean features found, particularly the sudden drop in energy loss to nearly the ionization value for $E \simeq 2{\cdot}5 \times 10^9$ e-volts, and the subsequent slow rise again for $E > 10^{10}$ e-volts are probably reliable indications of the behaviour of the penetrating component It must be remembered that the component is certainly complex, as it undoubtedly contains a fairly large fraction of protons. The fact that the energy at which the sudden fall of energy loss occurs is about equal to $2MC^2$, where M is the mass of a proton, suggests possible theories of the effect (Blackett 1936).

9·2. EVIDENCE FOR THE VALIDITY OF THE RADIATION FORMULA FOR HEAVY ELEMENTS

Some support for the validity of this formula for electrons in lead up to very high energies can be obtained by comparing the theoretical energy distribution giving the number of rays with energy greater than E, $g(E) \propto E^{-\gamma}$, where, as above, γ is of the order of 1·5, with that obtained from the observed distribution of bursts

Montgomery and Montgomery (1935*a*) have shown that the number of bursts with more than N rays under a 1 cm lead absorber, is proportional to N^{-s}, where s is about 2·2 Now from Table IV of the paper by Bhabha and Heitler (1937), it is seen that for $l = 3$ (1·2 cm Pb), the number of rays in a shower due to an electron of energy E, can be represented approximately by $N \propto E^q$, where $q \simeq 0{\cdot}8$ for $y = 5$ ($E = 10^9$ e-volts) and $q = 0{\cdot}4$, for $y = 10$ ($E = 10^{11}$ e-volts) The number of electrons with energy greater than E must therefore be proportional to E^{-qs} Since according to the cascade theory the number should be proportional to $E^{-1{\cdot}5}$, we should have $qs = 1{\cdot}5$ Actually, putting $s = 2{\cdot}2$, and q equal to 0·6, that is to the mean of the two values given above, we get $s = 1{\cdot}3$, in good agreement with the theory.

For thin absorbers (l small) we should expect from fig 4 of the paper by Bhabha and Heitler (1937) that q should be smaller than the figures given above; thus s should be larger This explanation is in agreement with the results of Montgomery and Montgomery (1935*b*), who find for a magnesium absorber a value of s considerably larger than 2·2.

I wish to express my gratitude to Mr J G. Wilson and Mr A. H Chapman for their invaluable help I am indebted also to the Government Grants Committee for a grant for apparatus. The work was carried out at Birkbeck College, London

SUMMARY

1 Further measurements of the energy loss of cosmic rays in metal plates have confirmed the former results of Blackett and Wilson (1937), that very nearly all the rays with $E < 2 \times 10^8$ e-volts are apparently electronic in character.

2 A reinterpretation of the data for higher energies, where the mean loss is quite small, in the light of the cascade theory of showers, has confirmed the conclusion of Neddermeyer and Anderson (1937) that there are probably present a few energetic, but absorbable, rays which can be identified with fully radiating electrons The penetrating rays, which at sea-level are in a large majority, cannot therefore be normal electrons

3 It is shown by indirect argument from the spectrum, and by direct experiment, that these penetrating rays become indistinguishable from electrons when their energy falls much below 2×10^8 e-volts.

4 It is pointed out that the main requirement of a theory of the penetrating component is to explain this striking property. Various possible theories are discussed, and it is shown that there are two main types of possible explanation The first is that the particles are heavy when energetic, but change their mass suddenly, during collisions with nuclear fields, perhaps through some type of resonance effect, when their energy falls below a critical energy. The second is that in which the penetrating rays are supposed to have the electronic rest mass, but are distinguishable from normal electrons by some unknown property which has the effect of making their radiative energy loss vary with their energy and the absorber, in a way rather similar to that given by the various theories previously put forward to explain the then supposed suppression of the radiation loss of normal electrons Further experiments will be required to distinguish between these possibilities.

REFERENCES

Bethe and Heitler 1934 *Proc Roy Soc* A, **146**, 83
Bhabha and Heitler 1937 *Proc. Roy Soc* A, **159**, 432
Blackett 1936 "Cosmic Rays" Oxford
— 1937*a* *Proc Roy. Soc* A, **159**, 1
— 1937*b* *Proc Roy Soc.* A, **159**, 19
Blackett and Wilson 1937 *Proc. Roy Soc.* A, **160**, 304.
Carlson and Oppenheimer 1937 *Phys Rev* **51**, 220
Crussard and Leprince Ringuet 1937 *C.R. Acad Sci , Paris*, **204**. 240.
Ehrenberg 1936 *Proc Roy Soc* A, **155**, 532.
Heitler 1937 *Proc Roy Soc* A, **161**, 261.
March 1937 *Z. Phys.* **108**, 128.

Montgomery and Montgomery 1935*a* *Phys. Rev* **48**, 786
— — 1935*b* *Phys Rev* **48**, 969.
Neddermeyer and Anderson 1937 *Phys. Rev* **51**, 884.
Nordheim 1936 *Phys Rev* **49**, 189
Street and Stevenson 1936 *Phys Rev* **49**, 425.
- 1937 *Phys. Rev* **52**, 1003
Turin and Crane 1937 *Phys Rev* **52**, 610.
Williams 1934 *Phys. Rev* **45**, 729
Wilson 1938 *Proc Roy Soc* (in the Press).

The continuous absorption spectra of alkyl iodides and alkyl bromides and their quantal interpretation

BY D. PORRET AND C F. GOODEVE

From the Sir William Ramsay and Ralph Forster Laboratory of Chemistry, University College, London, W C. 1

(*Communicated by F G Donnan, F R S.—Received 22 October* 1937)

In recent years a number of continuous absorption spectra of diatomic molecules have been examined in order to determine the potential energy-distance curves for the excited states. Gibson, Rice and Bayliss (1933) and Bayliss (1937) applied the methods of quantum mechanics to the chlorine and bromine spectra respectively and Stuckelberg (1932) treated the oxygen continuum in a similar manner The spectra of hydrogen bromide and iodide were analysed by Goodeve and Taylor (1935, 1936) The method of treatment has recently been extended and applied to the methyl bromide spectrum by Fink and Goodeve (1937).

In the present paper, recently published measurements for methyl iodide (Porret and Goodeve 1937) are analysed and compared with new data for ethyl iodide. The spectra of ethyl and butyl bromides have been measured and compared with that of methyl bromide.

1. THE QUANTAL INTERPRETATION OF THE METHYL IODIDE CONTINUUM

(*a*) *Outline of method*

The extinction coefficient, which is the probability of absorption of a quantum of light by a molecule, is determined by the eigenfunctions of the initial and final states and by the dipole moment associated with the transition. From known potential energy curves, it is easy to calculate the approximate extinction coefficients for the different frequencies, but *direct* calculation in the reverse direction is impossible in practice. To determine the potential energy curve for the upper state, one has first to calculate α, the absolute extinction coefficient, as a function of r, the internuclear distance, by quantal methods. By comparison with the experimental curve for α in terms of frequency one obtains the required relation between the potential energy (expressed as the frequency interval from the zero point energy level of the ground state, plus the zero point energy) and r.

The method of calculation of the α-r curve is described in detail by Fink and Goodeve (1937). α is given by the equation

$$\alpha = K\nu\left[\int \psi' M \psi''\right]^2, \tag{1}$$

where ν is the frequency, ψ' and ψ'' the eigenfunctions of the upper and lower states, M the dipole moment and K a constant. As the nuclei represent practically the whole mass of the atoms, the variation of the nuclear eigenfunctions is generally the determining factor in the value of the above integral All the other terms may, therefore, be assumed constant and one may write

$$\alpha = K_1\nu\left[\int_0^\infty \psi'_N \psi''_V dr\right]^2 \tag{2}$$

It is found that for alkyl halides, and for the range of extinction coefficients with which we are dealing, molecules possessing vibrational energy up as high as the third quantum level, play a part in the absorption. The eigenfunction curves of the levels of the ground state can be calculated by means of well-known equations The one of the upper state can be obtained according to the Wenzel-Kramers-Brillouin approximation from the solution of the wave equation

$$\frac{d^2\psi}{dx'^2} + \frac{8\pi^2\mu}{h^2} k'x'\psi = 0, \tag{3}$$

where k' is the slope of the upper potential energy curve, μ the reduced mass and x' the internuclear distance, defined here by $x' = r - r''_e$. The slope, k', can be calculated by means of the approximation used by Goodeve and Taylor (1935) which gives accurate values of the slope for small values of x' (see Fink and Goodeve 1937, section 4).

The value of K_1 can be determined by calculating α_0, the probability of transitions from the zero vibration level, multiplying this by the partition fraction for this level, f_0 (i.e. the fraction of the molecules in this level), and equating the maximum value of $\alpha_0 f_0$ to the maximum of the experimental extinction coefficient curve. For the region of the maximum the zero vibration level is solely responsible for the absorption. K_1 has the same value for all transitions between the two electronic states, so far as the assumptions in equation (2) are valid

For methyl iodide the variation of the frequency throughout the band is small and sufficiently accurate values for use in equation (1) can be obtained by application of the method of analysis used by Goodeve and Taylor (1935).

For any particular frequency, the observed extinction coefficient is the sum of the probabilities of transitions arising from the various vibrational levels in the ground state, each probability being multiplied by the appropriate vibrational partition fraction. Before making the summation one has to take account of the fact that the transitions do not involve the same frequency for a particular value of x' (see equation (17), Fink and Goodeve (1937)) Furthermore, the eigenfunctions of the ground state have been calculated for a harmonic oscillator. The formulae for an anharmonic one are very complicated, but a partial correction can be made by shifting each of the ψ'' curves so that its centre corresponds to the mid-point of the appropriate vibration level (see fig. 3) Details of the method of applying these corrections are given by Fink and Goodeve (1937) The corrections having been made, the total extinction curve can be calculated in terms of x' and, by comparison with the experimental one, the upper potential energy curve can be drawn.

(*b*) *Application to methyl iodide*

As with methyl bromide, the spectrum of the iodide can be considered as arising from a diatomic molecule (CH_3) –I The necessary data for methyl iodide are given in the table, section (4). The value of the slope, k', to be used in the solutions of equation (3) was found to be $-3{\cdot}8 \times 10^4$ wave-numbers per Å or $-7{\cdot}47 \times 10^{-4}$ ergs per cm. The eigenfunction curve for the level corresponding to $x' = 0$ is shown in fig 3 The integration of equation (2) was conveniently replaced by a summation and the procedure, as described above, carried out.

The partial extinction coefficients (as logarithms) for transitions arising from the first four vibration levels (at 20° C) are shown in fig. 1, plotted against the value of x'. The curve for the total extinction coefficient is also shown (curve I). Comparison of this curve with the observed extinction coefficient curve, shown by the full line in fig. 2, gives curve I in fig. 3 as the upper potential energy curve. In fig. 3 are also shown the lower potential energy curve and the first three vibration levels with their eigenfunction curves in arbitrary units.

It will be seen that this analysis leads to an upper potential energy curve with a marked point of inflexion, the repulsive force actually increasing

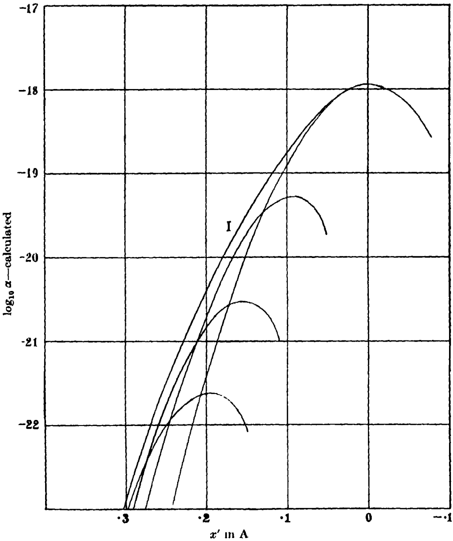


FIG 1. The extinction coefficient calculated as a function of the carbon-iodine distance

with increase of nuclear separation A close examination of the approximations used in the analyses has revealed no likely source of error which could account for this result. It may be that the absorption band is in reality a double one, i e. transitions to two separate upper states occur in the same

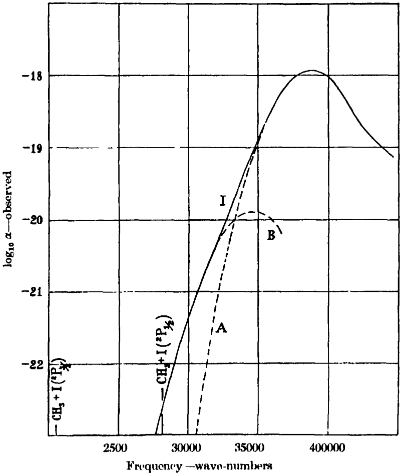


FIG. 2. The observed extinction coefficient curve of methyl iodide and the proposed resolution into two bands.

region of the spectrum The point of inflexion may in fact be removed by resolving the extinction coefficient curve into two parts, as shown by the broken lines in fig. 2. By comparing curve I in fig. 1 with the principal band, *A*, shown in fig. 2, an upper potential energy curve shown by the broken line, curve *A*, fig. 3, is obtained. It may be seen that this has the normal

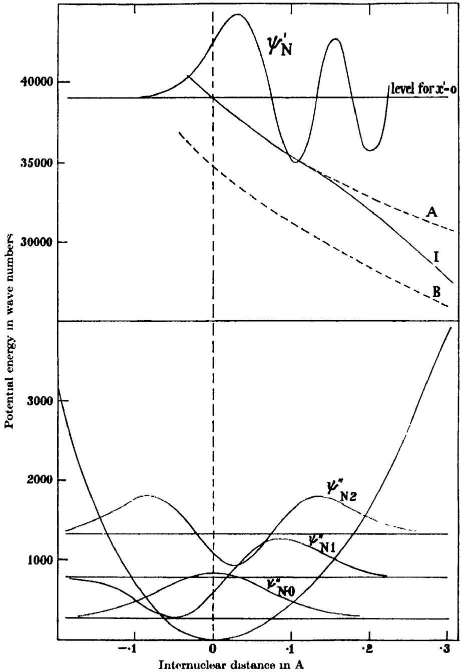


Fig. 3. The potential energy curves of the C-I link and the eigenfunction curves of some of the energy levels.

shape The approximate position of the upper potential energy curve corresponding to the *B* band is indicated by curve *B*, fig. 3.

The argument that leads to the conclusion that the band is double is easily seen by inspection of the two curves, I, fig 1, and I, fig 2. The former, the calculated curve, is curved throughout its whole length, whereas the latter, the experimental curve, has a marked straight part with a slight depression.

(*c*) *Influence of temperature*

Some observations were made of the extinction coefficient at 100° C The results were as follows.

Frequency cm $^{-1}$	$\log_{10}\alpha$, 20° C. Observed	$\log_{10}\alpha$, 100° C. Observed	$\log_{10}\alpha$, 100° C. Calculated
29,200	$2\bar{2}{\cdot}13$	$2\bar{2}{\cdot}47$	$2\bar{2}{\cdot}99$
29,800	$2\bar{2}{\cdot}51$	$22{\cdot}83$	$2\bar{1}{\cdot}31$
30,200	$2\bar{2}{\cdot}75$	$2\bar{1}{\cdot}03$	$21{\cdot}49$
31,000	$2\bar{1}{\cdot}23$	$2\bar{1}{\cdot}44$	$21{\cdot}83$

The values given in the last column are calculated assuming that the band is a single one, i.e. that the potential energy curve I, fig. 3, is the correct one. The observed extinction coefficients are about three times smaller than the calculated ones, from which it is to be concluded that the band is not single. Calculations based upon the resolution of the curves as shown in fig 2 yield results in agreement with the observations, but, as the observations were made at two temperatures only, this agreement cannot be considered as a quantitative confirmation of the resolution as shown.

(*d*) *Discussion*

The presence of two bands for methyl iodide brings its spectrum into conformity with those of the inorganic iodides (see Porret and Goodeve 1937). It would appear, therefore, that the *A* band leads to dissociation giving an excited $^2P_{\frac{1}{2}}$ iodine atom and a normal unexcited methyl free radical, whereas absorption of light in the *B* band yields products both of which are unexcited.

It is of interest to note that Mulliken (1935) discusses arguments which suggest that the observed continuum of methyl iodide is complex.

2. THE EXTINCTION COEFFICIENTS OF ETHYL IODIDE

The continuous absorption spectrum of ethyl iodide has been studied by a number of observers in order to obtain the position of the maximum or the "long wave absorption limit" (Scheibe 1925, Iredale and Wallace 1929, Iredale and Mills 1930, 1931, Emschwiller 1930, Hukumoto 1932; Scheibe, Povenz and Lindström 1933) According to Iredale and Mills (1930) and to Hukumoto (1932) the limit corresponds to a dissociation into a $—C_2H_5$ radical and a $^2P_{\frac{1}{2}}$ iodine atom. It is now well known that such a conclusion cannot be drawn, as the limit of the absorption shifts towards the red when the pressure or the length of the absorbing column is increased

Pure ethyl iodide supplied by the British Drug Houses, Ltd , was used after being fractionated three times and dried with P_2O_5. The measurements were carried out exactly as for methyl iodide (Porret and Goodeve 1937)

The measurements of the molar and absolute extinction coefficients, defined by the equations

$$\epsilon = \log_{10} \frac{I_0}{I_t} \cdot \frac{1}{c.l} \quad \text{and} \quad \alpha = \log_e \frac{I_0}{I_t} \cdot \frac{1}{n.l}, \tag{4}$$

respectively (l being the length in cm and c and n being the concentration expressed in g mol. per litre and in molecules per c c respectively), are plotted as their logarithms in fig. 4 The curve for methyl iodide is shown for comparison.

3 THE EXTINCTION COEFFICIENTS OF ETHYL AND BUTYL BROMIDES

Iredale and Mills (1931) have measured qualitatively the absorption of ethyl bromide, but no measurements of the absorption of the butyl compound have been recorded

Pure bromides supplied by the British Drug Houses, Ltd., were used and measurements made as for methyl iodide

The results for ethyl bromide are indicated in fig 4 by the open circles and the broken line and those for butyl bromide by the black circles and the dotted line The curve obtained by Fink and Goodeve (1937) for methyl bromide is shown for comparison

4 THE INFLUENCE OF THE CARBON CHAIN

It will be seen from fig 4 that the extinction coefficient curves are practically unaltered by an increase in the length of the carbon chain This, at first sight, is surprising, as the increased mass of the alkyl group might be expected to influence the shape of the band The coincidence between the curves, however, can be explained by a consideration of the structure of the molecules and of the fundamental vibrational frequencies.

At room temperature the methyl group can be treated as a unit of mass 15 because the vibrations involving movement of the hydrogen atoms with respect to the carbon are relatively unexcited. For ethyl and higher halides the alkyl group does not, however, behave as a unit This is indicated by the

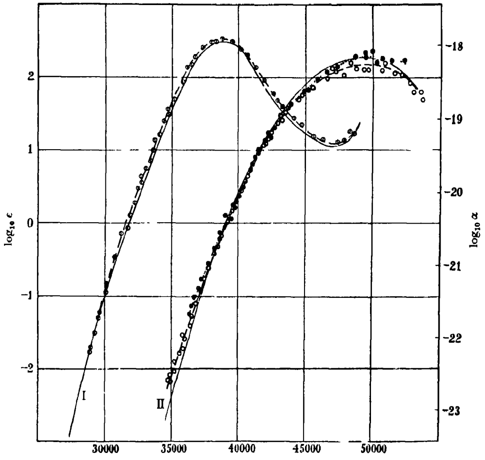


FIG. 4 Observed extinction coefficients ⦶ Ethyl iodide, continuous curve I, Methyl iodide; ○ Ethyl bromide, ● Butyl bromide, continuous curve II, Methyl bromide.

presence of low fundamental frequencies corresponding to vibrations in which the separate carbon atoms move individually towards the halogen atom (Kohlrausch 1931). Owing to the zigzag structure of the carbon chain and the low angular restoring forces the second and subsequent carbon atoms exert little restriction on the movement of the first. This is seen by an inspection of the carbon-halogen frequencies in a homologous series as set out in the table, column 2. This frequency changes only slightly on passing from methyl to ethyl halides and remains practically constant for the higher members of the series. If the alkyl group vibrated as a whole the variation would be much larger. Assuming that the force constant, F, of the carbon-halogen bond is the same for all the members of an homologous series, an "effective" reduced mass can be calculated from the fundamental frequency from the equation

$$\mu_\omega = \frac{F}{4\pi^2\omega^2}. \qquad (5)$$

The values given in the table, column 3, are obtained in this way and are the better ones to use in applying the above method of analysis to the ethyl and higher compounds as they take account of the slight restriction in the movement of the carbon atom of the carbon-halogen bond From the small variation in the values of μ_ω one would not expect much difference between the absorption curves of the various members of each series.

	ω	μ_ω	$x_e\omega$‖	D_e Kcal	f_0	f_1	f_2	f_3
CH_3I	533*	$2·21 \times 10^{-23}$§	3·50	106 ± 10	92·6	6·8	0·52	0·044
C_2H_5I	500†	2·51	3·05	—	91·5	7·8	0·67	0·058
C_3H_7I	503‡	2·48	—	—	—	—	—	—
C_4H_9I	505‡	2·46	—	—	—	—	—	—
CH_3Br	610*	2·08§	3·60	72 ± 10	94·85	4·9	0·26	0·015
C_2H_5Br	560†	2·47	3·05	—	93·6	6·0	0·39	0·025
C_3H_7Br	570‡	2·39	—	—	—	—	—	—
C_4H_9Br	557‡	2·50			—			

* Barker and Plyler (1935)
† Cross and Daniels (1933).
‡ Kohlrausch (1931)
§ Calculated directly, assuming CH_3— acts as a unit.
‖ Calculated from D_e (see Jevons 1932).

On the assumption that the upper and lower potential energy curves are unaffected, the influence of an increase of reduced mass on an absorption spectrum may be deduced. Such an increase or, in fact any restriction to motion, causes a "narrowing" of the eigenfunction curves, leading to a narrowing of the absorption band. At the same time, as a consequence of normalization, the height of the maximum is increased. For ethyl iodide the increase in α_{max} should be about 3 %. On the other hand the fundamental frequency ω is reduced, resulting in a lower value of the partition fraction, f_0, and a lower value of α_{max}. For ethyl iodide this lowering is about 1 %. These effects are smaller than the experimental error For the bromides the maxima occur in a region of high experimental error and the small difference shown in fig. 4 may not be a real one

The effect of an increase of the reduced mass on the lower part of the curve is more complex. The narrowing of the eigenfunction curves produces a narrowing of the absorption curve, but the increase in the partition fractions of the first and higher levels produces a broadening A small additional narrowing results from the correction given in equation (16) in the paper by Fink and Goodeve (1937) All these corrections are small. The small differences shown may be due to any one of these causes or to impurities.

5 CONCLUSIONS

The C-I and C-Br absorption bands differ in that the former is much narrower than the latter The upper potential energy curve associated with the C-I band is much less steep than that with the C-Br band, the values at $x' = 0$ being -4.3 and -7.1×10^4 wave-numbers per A respectively. This is the major cause of the difference in the shapes of the curves—the differences in the reduced masses and in the lower potential energy curves being small

The slope of the upper potential energy curve is equal to the force of repulsion The lower value for the C-I link is to be expected from the greater polarizability of the iodine atom and from the fact that the iodine atom is further from the carbon than is the bromine atom, the distances, r_{r_0}, being 2·12 and 1 88 A respectively (Sutherland 1937) The lower force of repulsion for the C-I link in the upper state is to be compared with the lower force constant of the link in the ground state

In the two homologous series of compounds discussed above, it is seen that the individuality of the chromophoric (light absorbing) grouping is retained in each series over a very wide range of extinction coefficients. It follows that the absorption process with these molecules is localized in one part. This is in contrast with what occurs in dyes or other mesomeric systems where the absorption process is bound up with the molecule as a whole. In the absence of mesomerism, localization of the absorption is to be expected, but the position of a band may be affected by the presence of groups which disturb the electronic levels of the chromophoric grouping

One of us (D. P.) wishes to express his thanks to the Swiss Ramsay Committee and the Ramsay Memorial Fellowship Trustees for the award of a Fellowship.

SUMMARY

The extinction coefficient curve for methyl iodide has been analysed according to the method used by Fink and Goodeve, and the existence of two bands has been deduced The upper potential energy curves corresponding to these bands have been determined.

The extinction coefficients of ethyl iodide and of ethyl and butyl bromides have been determined over a wide range. Influence of the carbon chain has been found to be small The conditions under which chromophoric groupings retain their individuality have been discussed.

REFERENCES

Barker, E F and Plyler, E. K 1935 *J. Chem. Phys.* **3**, 367–8.
Bayliss, N S. 1937 *Proc Roy Soc* A, **158**, 551–61
Cross, P. C and Daniels, F 1933 *J Chem Phys* **1**, 48.
Emschwiller, G 1930 *C R Acad Sci*, *Paris*, **191**, 208.
Fink, P. and Goodeve, C. F. 1937 *Proc Roy Soc* A, **163**, 592.
Gibson, G E, Rice, O K and Bayliss, N S 1933 *Phys Rev.* **44**, 193–200.
Goodeve, C. F and Taylor, A. W. C 1935 *Proc Roy Soc* A, **152**, 221–30.
— — 1936 *Proc Roy Soc* A, **154**, 181–7
Hukumoto, Y 1932 *Sci Rep Tôhoku Univ* **21**, 906–27.
Iredale and Mills 1930 *Nature, Lond*, **126**, 604
— — 1931 *Proc. Roy. Soc* A, **133**, 430.
Iredale and Wallace 1929 *Phil. Mag.* **8**, 1093
Jevons, W 1932 "Report on Band Spectra of Diatomic Molecules", p 27. Physical Society, London
Kohlrausch, K W F 1931 "Der Smekal-Raman Effect", p. 146
Mulliken, R S. 1935 *Phys Rev* ii, **47**, 413–15
Porret, D and Goodeve, C F. 1937 *Trans Faraday Soc.* **33**, 690–3.
Scheibe, G 1925 *Ber dtsch chem Ges* **58**, 586
Scheibe, G, Povenz, F and Linstrom, C F. 1933 *Z phys. Chem* B, **20**, 283.
Stuckelberg 1932 *Phys Rev* **42**, 518
Sutherland, G B B. M 1937 *Nature, Lond*, **140**, 239

The infra-red absorption spectrum of methylene chloride

By C Corin* and G. B B. M Sutherland

(*Communicated by R G. W. Norrish, F R S* -*Received* 29 *October* 1937)

The absorption spectrum of methylene chloride has been studied over limited regions in the infra-red by several workers (Ellis 1926, Lecomte 1933, Corin 1936, Emschwiller and Lecomte 1937) So far as we are aware it has never been investigated between 2·5 and 7μ, although this is a very important region The present paper describes a study of the absorption of this substance in the liquid state between 2 and 12μ, which was undertaken with the following objects in view First, it was hoped that it might be possible to resolve certain difficulties in the correlation of the infra-red with the Raman spectrum of this molecule Thus Trumpy (1934) from a very careful examination of the Raman spectrum had made an assignment of the nine fundamental frequencies which did not admit of a reasonable explanation of a very strong absorption observed in the infra-red by Lecomte (1933) near 8μ Before any reinterpretation of existing data could be attempted it was obviously necessary to complete the map of the spectrum between 2·5 and 7μ and to confirm Lecomte's work at 8μ. It was also important to observe as many overtone and combination bands as possible to provide data for verifying and testing any new assignment of the fundamental frequencies Secondly, it was chosen as a suitable molecule for the further study of the assumption of independent groups (Sutherland and Dennison 1935, Mecke 1932) in computing the frequencies of a polyatomic molecule from a simplified but still very general potential function. Thus it is well known that certain groups (such as CH_2 or CH_3) in a molecule give rise to certain characteristic frequencies in the infra-red and Raman spectra of these molecules, being apparently practically independent of the rest of the molecule Such frequencies are easy to identify in analysing the vibration spectrum of a polyatomic molecule but the frequencies which arise from the vibrations of those semi-independent groups as wholes are not at all easy to assign and it is just those "inter-group" frequencies as opposed to the "intra-group" frequencies for which we have found a new method of treatment in this case Yet the method, which depends on an application of the isotope effect, is a very general one and should be applicable to many classes of polyatomic molecules.

---

* Aspirant du Fonds National de la Recherche Scientifique.

EXPERIMENTAL

The spectrometer used was the well-known Hilger instrument D83. The source of radiation was a Nernst glower, while the currents from the thermopile were measured directly on a very sensitive Paschen type galvanometer The spectrometer was carefully enclosed in a wooden box to minimize thermal disturbances The dispersing prism was of fluorite between 2 and 9μ, and of rock salt between 8 and 12μ. The wave-length readings on the drums were checked with reference to the well-known atmospheric bands due to water vapour and carbon dioxide, the prisms being set in position initially by using the sodium *D* lines No special precautions were taken to prevent variations in the current through the Nernst filament since the latter was operated from a battery supply which was very constant. For the majority of measurements the Paschen galvanometer mentioned above was employed, occasionally, however, the thermo-electric current was taken to a Kipp and Zonen Zc moving coil galvanometer the deflexions of which were amplified by a photo-electric relay of the type described by Moss (1935) This made it possible to work with very fine slits in the examination of the contours of certain bands.

The absorption cell employed is shown diagrammatically in fig 1. It enabled one to vary the thickness of the layer of CH_2Cl_2 from 0·007 to 1 mm The liquid is contained in the space *A* between the rock salt windows *B* and *C*, which were 5 mm. thick and 20 mm. in diameter * These windows are surrounded by two rings, *D* and *E*, the latter being separated by a third ring, *F*, the thickness of which determines that of the layer of liquid to be examined. The ring *D* is provided with two tubes *G* and *H* which allow the liquid to flow into the cell through the openings *I* and *J*. The ring *F* (the internal diameter of which is a shade less than that of *D* or *E*) is placed against the ring *D* so that the openings *I'* and *J'* of the former correspond to *I* and *J* of the latter The ring *E* is identical with *D* except that it has no tubes. These three rings are held together by screws through K_1, K_2, K_3 and K_4. The rock salt plates are next placed on either side of *F* and the whole pressed together tightly by means of two larger metal rings *L* and *M*, lead washers being put between *D* and *L* on one side and *E* and *M* on the other The thickness of *D* (and of *E*) was just a little less than that of the rock salt plate. The cell was filled through either *G* or *H*, one being the entrance tube, the other the exit for the air. This method proved very satisfactory even for the smallest thicknesses. The whole was made of brass, except for the washer *F* which was made from brass, aluminium or copper according to the thickness required

---

* We are much indebted to Professor Errera of the University of Brussels for the loan of these rock salt plates.

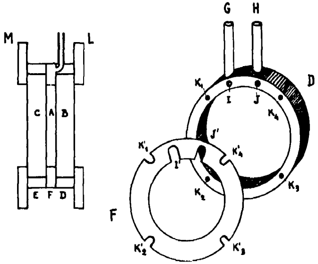


Fig. 1. The absorption cell.

In making the observations a preliminary run was done with the empty cell to ensure that the windows were perfectly clean and to get an estimate of the percentage loss due solely to them. This was of the order of 15 %. Readings were then taken of the energy falling on the thermopile with, and without, the cell in the path of the beam to give the percentage absorption at any particular wave-length Beyond 4μ a glass shutter was always used for cutting off the radiation from the thermopile to ensure that only the long wave-length radiation was being measured. The thickness of the cell varied from place to place in the spectrum depending on the intensity of the absorption at the particular region It was chosen to yield a maximum absorption at the centre of the band of from 60 to 90 %. All the important bands were plotted on at least two independent occasions and indications of weak bands were confirmed or disproved by using thicker layers of liquid. The width of the entrance and exit slits were always the same and were always as narrow as would still yield reasonable deflexions They are indicated to scale for each region on its appropriate diagram.

RESULTS

Between 2 and 12μ we have observed altogether twenty-six absorption bands of which sixteen (occurring between 2·9 and 7μ) are new, the remaining ten having been already reported by earlier workers The appearance of these bands is illustrated in figs 2 and 3. For convenience of reference the various bands have been numbered beginning from those of highest frequency The thickness of the layer used is indicated on each curve in mm together with the slit width. The percentage absorptions plotted are actually those observed and have not been corrected for absorption and scattering due to the cell alone.

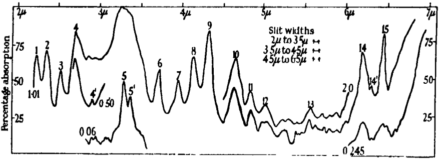


FIG. 2. Spectrum of methylene chloride from 2 to 6 5μ.

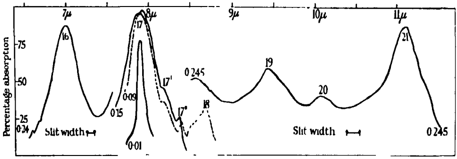


FIG. 3. Spectrum of methylene chloride from 6 5 to 11·5μ.

It will be observed that the most intense absorption bands occur at 3·3, 7, 7·9 and 11·1μ, bands of medium intensity were found at 4·3, 6·3, 6·5 and 9·4μ, while seventeen very much weaker bands were observed at the positions given in Table I The band at 3·3μ was found on using a smaller thickness of liquid and very narrow slits to consists of two bands having their maxima at 3043 and 2985 cm $^{-1}$. These observations are compared with the existing data on the infra-red and Raman spectrum of this molecule in Table I. It will be noticed that in cases where the observations overlap the agreement between the positions of the centres of the bands as reported by the different authors is quite satisfactory, with the exception of the weak bands near 8·6 and 9·5μ. Even in these two cases it probably does not exceed the experimental error, since the bands in question are rather broad and do not have a well-defined maximum. The coincidence between the two strong Raman lines lying at 3043 and 2985 cm^{-1} and the strong absorptions at 3049 and 2985 cm^{-1} is especially satisfactory. The most remarkable fact is that the strongest absorption band of all, viz. that at 7·9μ, has no counterpart in the Raman spectrum. It is this which makes necessary the detailed discussion of the fundamental frequencies of the CH_2Cl_2 molecule in the section which follows.

TABLE I. THE VIBRATION SPECTRUM OF METHYLENE CHLORIDE

	Present investigation			Lecomte	Ellis	Corin	Raman spectrum
	Position						
	In μ	In cm.$^{-1}$	Intensity	in μ	in μ	in μ	in cm.$^{-1}$
1	2·24	4464	Weak	-	2·27	2·24	—
2	2·36	4237	Weak		2·42	2·33	—
3	2·54	3937	Weak		2·57	2·57	—
4	2·70	3704	Weak	—	2·76	—	
4′	2·92	3425	Very weak	—	2·80	—	
5	3·28	3049	Strong	—		—	3043
5′	3·35	2985	Strong	- -	—		2985
6	3·74	2674	Weak		-		—
7	3·96	2525	Weak	-	-	—	- -
8	4·14	2415	Weak	—			-
9	4·32	2315	Medium		-		—
10	4·68	2137	Weak		-	-	—
11	4·86	2058	Weak	—	-	—	-
12	5·04	1984	Very weak		—		—
13	5·58	1792	Very weak	-		-	—
14	6·20	1613	Weak	- -		—	-
14′	6·32	1582	Very weak		—		—
15	6·46	1548	Weak		-	-	—
16	7·00	1429	Strong	7·00			1419
17	7·90	1266	Very strong	8·02			-
17′	8·18	1222	Weak				-
17″	8·39	1192	Weak	—		—	
18	8·66	1155	Weak	8·83			1149
19	9·43	1060	Weak	9·59			1060
20	10·14	985	Weak	10·28		—	-
21	11·12	899	Medium	11·20	-		898
22	—	737	Very strong	13·60	-	-	735
23	—	704	Very strong	14·20		- -	700
24	Not observed in infra-red up to the present						284

DISCUSSION

The molecule CH_2Cl_2 has nine fundamental frequencies. These may be separated into three groups of three in a rough physical way by considering the molecule to be built up from a CH_2 group and a CCl_2 group. If we look at the molecule in this way then it is clear that the CH_2 group and the CCl_2 group will have each three fundamental frequencies, while the vibrations of the CH_2 group as a whole with respect to the CCl_2 group as a whole form the third group of three (Group 3) This method of splitting the molecule into semi-independent vibrating groups has been shown to be justified by Sutherland and Dennison (1935) and by other authors (Mecke 1932, Kohlrausch 1935) Its great value is that it enables some of the fundamental frequencies to be immediately identified. Thus the CH_2 group is well known to possess three fundamental frequencies, two of which fall near 3000 cm $^{-1}$ and the other near 1450 cm.$^{-1}$. We are therefore fairly sure that the bands at 3049, 2985 and 1430 cm.$^{-1}$ are the first group of three fundamentals In a similar way one may pick out the three frequencies at 700, 735 and 284 cm $^{-1}$ as the second group of three, consisting essentially of the vibrations of the CCl_2 group. The real difficulty comes when we try to say which are the other three fundamentals When the Raman spectrum of CH_2Cl_2 was examined in great detail by Trumpy (1934) he found, in addition to the above six fundamentals (which he assigned in the same way as we have done), three other Raman lines at 898, 1060 and 1149 cm $^{-1}$ Since all of the nine fundamental frequencies of this molecule are permitted in the Raman effect, Trumpy concluded very naturally that these frequencies represented the other three fundamentals. The objection to this assignment is that it leaves quite unexplained the most intense infra-red absorption at 1266 cm.$^{-1}$. This absorption is so strong that it can scarcely be considered to be other than a fundamental frequency. Even if one attempts to interpret it as a combination band of Trumpy's fundamentals one meets with very serious difficulties since the only combination which it might be is $\nu_1+2\delta_1$.* That such a combination band should have powerful absorption is extremely unlikely and we propose to accept the strong infra-red frequency at 1266 cm $^{-1}$ as one of the three frequencies of Group 3. The problem is now to decide the other two.

We first notice that of the three Raman frequencies of Trumpy at 898, 1060 and 1149 cm.$^{-1}$ only one appears with any marked intensity in the infra-red this is the one at 898 cm.$^{-1}$. If we accept this as the second of the missing fundamentals then it remains to decide between 1060 and 1149 cm. ^{1} as the last. One must not neglect the possibility that both of those frequencies (being relatively weak in absorption and in scattering) may be overtone or combination frequencies due to a low fundamental in the neighbourhood of 500 600 cm $^{-1}$ and Emschwiller and Lecomte (1937) have in fact made a suggestion of this kind. Since this region of the spectrum has not been investigated in the infra-red and was beyond the range of our spectrometer we have tackled the problem from the theoretical side and have been able to show that the possibility of such a low fundamental is most unlikely (see next section). If we rule out Lecomte's suggestion, we consider that the frequency at 1060 cm ^{1} is the correct one to choose as the remaining fundamental for the following reasons First of all, it is quite impossible to interpret it in terms of any simple combination of the other eight frequencies, secondly, it is possible to interpret the frequency at 1149 cm $^{-1}$ in terms of the other eight frequencies in at least two simple and reasonable ways, finally, we have been able to estimate the approximate value of this frequency as somewhere between 910 and 1110 cm $^{-1}$ by a calculation which we shall now describe.

---

* See Table III for explanation of this notation.

THE ISOTOPIC METHOD

The mathematical treatment of the normal vibrations of a system of the type YX_2Z_2 has not yet been given, and although it would be possible to extend the treatments given of simpler molecules to this type, the process would be extremely laborious and of no small difficulty since it would involve the solving of a ninth order determinantal equation The results obtained would also be dependent on the type of potential function assumed so that at best one might only obtain a somewhat forced agreement between calculated and observed frequencies It occurred to us that a relatively simple approximate treatment might be made to yield rough values for the frequencies by regarding the Cl atoms as heavy "isotopes" of hydrogen of mass 35 and then employ the equations for the isotope effect in CH_4 which have been derived by Rosenthal (1934) Of course the values one obtains for the three frequencies will be too high since the C-Cl force constant is appreciably smaller than the C-H force constant which has been taken in its place To get over this difficulty we have also calculated the same three frequencies looking on the hydrogen atoms as light "isotopes" of the Cl atoms and using the potential constants of the CCl_4 molecule given by Vogo and Rosenthal (1936) This will give too low a value for the frequencies since the C-Cl force constant has been used for two bonds in which the real force constant is somewhat higher We hoped that the difference between the upper and lower limits thus obtained would not be too great so that a correlation might be made with the observational data Table II shows that the experiment has been more successful than one might have imagined Thus, using Rosenthal's notation, we find that the highest of the three (ω_{4a}) should lie between 1295 and 1081 cm.$^{-1}$ This we identify with the strong infra-red band at 1266 cm^{-1}. The lowest (ω_{4c}) is to be expected between 720 and 843 cm^{-1}, this we associate with the frequency at 898 cm The middle one (ω_{2a}) should appear between 910 and 1108 cm^{-1} and would therefore correspond more closely to the frequency at 1060 cm^{-1} than to that at 1150 cm.$^{-1}$ While this evidence for the assignment of the three group frequencies could not be regarded as conclusive in itself, when taken into consideration with the remarks above concerning intensities, it would appear to be the most consistent correlation of the spectrum yet offered.

TABLE II THE METHYLENE CHLORIDE FREQUENCIES OF GROUP 3

Designation		Calculated values		
Rosenthal's notation	Present notation	Upper limit	Lower limit	Experimental values
ω_{4a}	ω_3	1295	1081	1266
ω_{2a}	ω_2	1108	910	1060
ω_{4c}	ω_1	843	720	896

It remains to show that the other bands can be interpreted in terms of the nine fundamentals we have now chosen That this can readily be done is shown in Table III. The main difficulty is to account for the frequency at 1150 cm^{-1}. This would appear to be either the addition band $\omega_1+\delta_1$ or the difference band $\delta_2-\delta_1$ cm or possibly a superposition of both of them, since the band appears to have two maxima in the infra-red and in the Raman spectrum there seems to be more uncertainty about its exact magnitude than is the case with the other frequencies Why this should be the only combination band to appear in the Raman spectrum is not at all clear. All that we can say here is that a similar phenomenon has been observed for other polyatomic molecules (e.g ethylene), and that the conditions determining the appearance of overtone and combination frequencies in Raman spectra are not yet understood As regards the other assignments in Table III we might remark that all of the bands with the exception of 4′ and 8 have been accounted for as simple combinations of not more than two fundamentals Since such simple addition bands are those which are most likely to appear, our assignment of the fundamentals appears to gain additional evidence. Our attempts to account for all of our observed bands in terms of Trumpy's fundamentals had not an equal measure of success In several cases we have given only one assignment where alternative interpretations were possible For instance, band 11 might equally well be interpreted as $2\omega_2$ and we have no reason for preferring the assignment given in Table III unless that it gives a slightly

TABLE III. OVERTONE AND COMBINATION BANDS OF METHYLENE CHLORIDE

Fundamentals arising from the CCl_2 group $\begin{cases} \nu_1 = 700 \\ \nu_2 = 736 \\ \delta_1 = 284 \end{cases}$ Fundamentals arising from the CH_2 group $\begin{cases} \nu_3 = 2985 \\ \nu_4 = 3046 \\ \delta_2 = 1425 \end{cases}$ Fundamentals of Group 3 $\begin{cases} \omega_1 = 896 \\ \omega_2 = 1060 \\ \omega_3 = 1266 \end{cases}$

Observed frequency		Assignment	Predicted frequency
20	985	$\nu_1+\delta_1$	984
18	1155	$\nu_4-\delta_1$	1141
17″	1192	$\omega_1+\delta_1$	1180
17′	1222	$\nu_1+2\delta_1$	1268
15	1548	$\omega_3+\delta_1$	1550
14′	1582	$\nu_1+\omega_1$	1596
14	1613	$\nu_2+\omega_1$	1632
13	1792	$\nu_2+\omega_2$	1796
12	1984	$\nu_2+\omega_3$	2002
11	2058	$2\omega_1+\delta_1$	2076
10	2137	$\omega_3+\omega_1$	2162
9	2315	$\delta_2+\omega_1$	2321
8	2415	$\omega_3+\omega_1+\delta_1$	2421
7	2525	$2\omega_3$	2532
6	2674	$\delta_2+\omega_3$	2691
4′	3425	$\nu_3+\nu_2-\delta_1$	3420
4	3704	$\nu_3+\nu_2$	3721
3	3937	$\nu_4+\omega_1$	3942
2	1237	$\nu_3+\omega_3$	4251
1	4464	$\nu_4+\delta_2$	4471

All frequencies are given in cm. $^{-1}$.

better numerical agreement with the observed frequency. This is not a sufficiently good reason, however, for excluding one assignment in favour of another, since exact numerical agreement is not to be expected because of anharmonicity and interactions. We have not given all the possible ways of interpreting each band because we were only interested in showing that all the observed bands could be accounted for very easily in terms of our fundamentals In actual fact there are only another two simple alternative assignments in addition to the one just quoted, they are.

Band 13 as $2\omega_1$,
Band 10 as $\nu_2+\delta_2$.

It will be noticed that the numerical agreement between an observed and a predicted frequency is within 1 or 2 %, except the band 17′ the exact position of which is impossible to determine since it appears as a shoulder on the very intense band 17.

ACKNOWLEDGEMENTS

One of us (C C) would like to express his gratitude for the patronage of the University of Liége which enabled him to carry out this work in the Physical Chemical Laboratory of the University of Cambridge. We are both grateful to Professor Norrish for his interest in this work and to Mr G. K. T. Conn who gave much valuable help with the experimental side of the work in its more intractable periods.

SUMMARY

The infra-red absorption spectrum of methylene chloride has been investigated in the liquid state between 2 and 12μ with a prism spectrometer Twenty-six bands have been observed of which sixteen have not been recorded before, the positions and intensities of the remaining ten agree well with the works of other observers The most important fact which emerges is that one of the very intense absorption bands has no counterpart in the Raman spectrum of this molecule. This has necessitated a new assignment of the fundamental frequencies, which has been done partly by the method of independent groups and partly by applying the theory of the isotope effect in a molecule of the type YX_4 The latter is a new and surprisingly successful method of forming a rough numerical estimate of the magnitude of certain of the frequencies of the YX_2Z_2 molecule from a knowledge of the potential constants of the YX_4 and YZ_4 molecules The twenty-six observed bands are accounted for very simply in terms of the new set of fundamental frequencies.

REFERENCES

Corin 1936 *J. Chim. phys* **33**, 448.
Ellis 1926 *Phys. Rev* **28**, 25
Emschwiller and Lecomte 1937 *J. Phys* **7**, 130.

Kohlrausch 1935 *Z. Phys. Chem* B, **28**, 340
Lecomte 1933 *C.R. Acad. Sci, Paris*, **196**, 1011.
— 1936 "Traite de Chimie Organique", **2**, 196. Paris Masson et Cie
Mecke 1932 *Z Phys Chem* B, **17**, 1.
Moss 1935 *Rev Sci Instr.* **12**, 141
Rosenthal 1934 *Phys Rev* **45**, 538, **46**, 730
Sutherland and Dennison 1935 *Proc Roy. Soc* A, **148**, 250
Trumpy 1934 *Z Phys.* **88**, 226
Voge and Rosenthal 1936 *J. Chem. Phys* **4**, 137.

On the velocity and temperature distributions in the turbulent wake behind a heated body of revolution

By S Tomotika, D Sc

Professor of Mechanics, Osaka Imperial University, Osaka, Japan

(*Communicated by G. I. Taylor, F R S —Received* 3 *November* 1937)

Introduction

1 The calculation of the distribution of mean velocity in the turbulent wake behind a body of revolution in a uniform stream of an incompressible fluid was first carried out by Swain (1929) by using Prandtl's momentum transport theory of turbulent motion The uniform stream was considered to be parallel to the axis of revolution of the body and the mean motion was assumed to be symmetrical about the same axis Swain further adopted Prandtl's assumptions that for sufficiently high Reynolds numbers and at a sufficient distance downstream, there is geometrical and mechanical similarity in different sections of the wake, and that the values of the mixing length, l, at corresponding points in different sections are proportional to the breadths of the sections. She also assumed, as done by Schlichting (1930) in his discussion of the two-dimensional wind-shadow problem, that the mixing length l is constant over any one section of the wake.

Now, taking the x-axis along the axis of revolution of the body and the r-axis perpendicular to it, let the undisturbed velocity be denoted by U. Then, the x-component of velocity, u, in the wake may be written as

$$u = U - u^*. \tag{1}$$

Swain's analysis shows that the distribution of u^* for the first approximation may be expressed in the form

$$\frac{u^*}{u_0^*} = (1-\xi^{\frac{3}{2}})^2, \tag{2}$$

where u_0^* is the value of u^* at the centre of the wake in any given section and $\xi = r/r_0$, r_0 being the value of r at the edge of the wake.

The distribution of temperature in the wake behind a body of revolution when the obstacle had been heated was not discussed by Swain, but according to the momentum transport theory of turbulence the distribution of temperature across the wake is identical with that of velocity. Thus, if θ denotes the difference in temperature between any point in the wake and that in the main stream and also if θ_0 denotes the value of θ at the centre of the wake, we have

$$\frac{\theta}{\theta_0} = (1-\xi^{\frac{3}{2}})^2. \tag{3}$$

2. In a recent paper†, Goldstein (1935) has made an application of Taylor's vorticity transport theory of turbulence to the calculation of the velocity and temperature distributions in the turbulent wake behind a heated body of revolution which is placed, as in Swain's case, in a uniform stream in such a way that the axis of the body is parallel to the undisturbed velocity.

Goldstein has obtained the equations of motion in cylindrical co-ordinates according to the vorticity transport theory on the assumption that not only the mean motion, but also the turbulent motion, is symmetrical about the axis of revolution of the body, so that the vortex lines are circles about the axis.

Then, it has been shown that if, as in Swain's case, the states of affairs in different sections at a sufficient distance downstream behind the body are assumed to be geometrically and mechanically similar, and also if the values of the mixing length l at corresponding points in different sections are supposed proportional to the radii of the sections, there is no real solution at all when the value of l is constant over any one section of the wake.

Further, it has been shown by Goldstein that if $l^2 \propto r^{-p}$, there are real solutions provided p is greater than unity. However, the assumption $l^2 \propto r^{-p}$ $(p \geqq 1)$ makes l infinite at $r = 0$, i.e. on the axis of symmetry, so that it is open to criticism. Moreover, the results of Fage's and Townend's observations (1932 *a*, *b*; 1936) on the turbulent motion in rectangular pipes as well as in circular pipes show that the type of turbulent motion assumed by Goldstein is unlikely to occur in reality. Thus, Goldstein's results are of rather small interest from the practical point of view

---

† "On the velocity and temperature distributions in the turbulent wake behind a heated body of revolution" (unpublished).

In the present paper an attempt is made to apply Taylor's modified vorticity transport theory of turbulence to the calculation of both the velocity distribution and the temperature distribution in the turbulent wake behind a heated body of revolution which is placed in a uniform stream such that its axis of revolution is parallel to the direction of the undisturbed velocity. The geometrical and mechanical similarity is assumed, with Prandtl and Goldstein, for the states of affairs in different sections at a sufficient distance downstream behind the body and also the isotropy in turbulence is assumed in accordance with the results of observations The theoretical curve of velocity distribution is compared with the observations of Schlichting and of Simmons.

The writer wishes to express his cordial thanks to Professor G. I Taylor, F.R S , for suggesting the present problem The calculations were mostly carried out in 1935 when the writer was working in Cambridge, England.

MODIFIED VORTICITY TRANSPORT THEORY

3. The modified vorticity transport-theory of turbulence put forward by Taylor (1935, 1937) is based on the assumption that the components of vorticity are transferred unchanged by turbulence in the same way that momentum is transferred according to Prandtl's momentum transport theory. A portion of the fluid is conceived to leave a certain position with the vorticity components of the mean motion and to retain those components till it mixes with its surroundings after traversing the mixing length.

Taylor has shown that, according to this modified vorticity transport theory, the equations of motion in rectangular co-ordinates are

$$X-\frac{1}{\rho}\frac{\partial p}{\partial x}=u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}+\frac{1}{2}\frac{\partial}{\partial x}(\overline{q'^2})+(\overline{w'\eta'}-\overline{v'\zeta'}), \tag{4}$$

where

$$\overline{w'\eta'}-\overline{v'\zeta'}=-\overline{(x-a)w'}\frac{\partial\eta}{\partial x}-\overline{(y-b)w'}\frac{\partial\eta}{\partial y}-\overline{(z-c)w'}\frac{\partial\eta}{\partial z}$$
$$+\overline{(x-a)v'}\frac{\partial\zeta}{\partial x}+\overline{(y-b)v'}\frac{\partial\zeta}{\partial y}+\overline{(z-c)v'}\frac{\partial\zeta}{\partial z}, \tag{5}$$

together with the equations formed by cyclic permutation.

In (4) (X, Y, Z) are the components of body force, ρ is the density of the fluid concerned which is assumed to be incompressible and p is the average value of the difference of the pressure from the hydrostatic pressure. (u, v, w) are the velocity components of the mean motion, while (u', v', w') are those of turbulence and $q'^2 = u'^2 + v'^2 + w'^2$. Also, (ξ, η, ζ) are vorticity components of the mean motion, so that

$$\xi = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}, \quad \eta = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}, \quad \zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y},$$

and (ξ', η', ζ') are the components of the turbulent vorticity.

Further, the equation of continuity is

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{6}$$

4 When the mean motion is symmetrical about an axis, as in the case of the mean motion in the turbulent wake behind a body of revolution which is going to be discussed in the present paper, it will be convenient to transform the equations of motion into cylindrical co-ordinates.

Let (x, r, θ) be cylindrical co-ordinates Assuming the mean motion to be symmetrical about the x-axis, the x- and r-components of velocity of the mean motion will now be denoted by u and v respectively without causing any confusion. Also we denote by (u', v', w') respectively the x-, r- and θ-components of velocity of the turbulent motion which is not assumed to be symmetrical about the x-axis, but is rather assumed to be isotropic at a sufficient distance downstream so that $\overline{u'^2} = \overline{v'^2} = \overline{w'^2}$

Then, the equations of motion in cylindrical co-ordinates of the modified vorticity transport theory for the mean motion symmetrical about the x-axis are obtained by direct transformations from (4), together with (5), and other similar equations When the fluid is free from body force, we have

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial r} = -\frac{\partial}{\partial x}\left(\frac{p}{\rho} + \tfrac{1}{2}q'^2\right) - \overline{l_x v'}\left(\frac{\partial^2 v}{\partial x^2} - \frac{\partial^2 u}{\partial x \partial r}\right)$$
$$+ \overline{l_r v'}\left(\frac{\partial^2 u}{\partial r^2} - \frac{\partial^2 v}{\partial x \partial r}\right) + \overline{l_\theta w'}\left(\frac{1}{r}\frac{\partial u}{\partial r} - \frac{1}{r}\frac{\partial v}{\partial x}\right), \tag{7}$$

together with one similar equation for the γ-direction, where l_x, l_r, l_θ are the components of the mixing length l in the directions of the x-, r- and θ-axes respectively. Also, p has the same meaning as before and

$$q'^2 = u'^2 + v'^2 + w'^2.$$

The equation of continuity in cylindrical co-ordinates is

$$\frac{\partial}{\partial x}(ru) + \frac{\partial}{\partial r}(rv) = 0. \tag{8}$$

CALCULATION OF VELOCITY DISTRIBUTION IN THE WAKE BEHIND A BODY OF REVOLUTION

5 We consider a body of revolution in a stream whose undisturbed uniform velocity is U parallel to the axis of revolution. If we assume that the mean motion in the turbulent wake behind the body is symmetrical about the axis, the equations of motion according to the modified vorticity transport theory are given by (7), together with one similar equation. The turbulent motion, however, is not assumed to be symmetrical about the axis, but is rather assumed to be isotropic at a sufficient distance downstream.

For a first approximation, at a large distance downstream in the wake we put

$$u = U - u^*. \tag{9}$$

We assume that u^*/U is small, and we also make the usual assumptions of the boundary layer theory that v and its derivatives are small in comparison with u and its corresponding derivatives, and that derivatives with respect to x are small compared with the corresponding derivatives with respect to r. Further, we assume that the derivatives of $(p/\rho + \frac{1}{2}q'^2)$ may be neglected for a first approximation.

Then, under these simplifying assumptions the equation of motion (7) now gives the following equation for u^* for a first approximation.

$$U\frac{\partial u^*}{\partial x} = l_r v' \frac{\partial^2 u^*}{\partial r^2} + l_\theta w' \frac{1}{r}\frac{\partial u^*}{\partial r}. \tag{10}$$

If the mixing length is small and the turbulence is isotropic, as assumed in the present paper, we have

$$\overline{l_r v'} = \overline{l_\theta w'}. \tag{11}$$

Following Prandtl, we put

$$l_r v' = l_\theta \overline{w'} = l^2 \left|\frac{\partial u^*}{\partial r}\right|, \tag{12}$$

where l is the mixing length

Then, remembering that u^* decreases outwards from the axis, we get ultimately

$$U\frac{\partial u^*}{\partial x} = -l^2 \frac{\partial u^*}{\partial r}\left(\frac{\partial^2 u^*}{\partial r^2} + \frac{1}{r}\frac{\partial u^*}{\partial r}\right). \tag{13}$$

6 Now, if D denotes the total drag on the body under discussion, it can be shown without difficulty that

$$D = 2\pi\rho U\int u^* r\,dr. \tag{14}$$

If we assume that at corresponding points of different sections r varies as x^m and u^* varies as x^{-n}, it follows from (14) that $n = 2m$. If then we put

$$\frac{u^*}{U} = \frac{1}{x^{2m}} f\left(\frac{r}{x^m}\right), \tag{15}$$

then in order that both sides of (13) should be of the same order of magnitude, l^2 must vary as x^{5m-1} for a given value of r/x^m. Further, if the value of l for a given value of r/x^m varies as the breadth of the wake, we must have $5m-1$ equal to $2m$, so that $m = \frac{1}{3}$.

Thus, if now we put

$$\eta = \frac{r}{x^{\frac{1}{3}}}, \tag{16}$$

$$\frac{u^*}{U} = \frac{1}{x^{\frac{2}{3}}} f(\eta), \tag{17}$$

and

$$l^2 = \phi(\eta)\, x^{\frac{2}{3}}, \tag{18}$$

equation (13) becomes

$$\frac{1}{3}\left(2f + \eta\frac{df}{d\eta}\right) = \phi(\eta)\frac{df}{d\eta}\left(\frac{d^2f}{d\eta^2} + \frac{1}{\eta}\frac{df}{d\eta}\right). \tag{19}$$

On the axis, u^* is positive and $\partial u^*/\partial r$ is zero. Therefore

$$(f)_{\eta=0} > 0, \quad \left(\frac{df}{d\eta}\right)_{\eta=0} = 0. \tag{20}$$

If we make the assumption that the value of the mixing length l is constant over any one section, we may write $\phi(\eta) = c$, where c is a certain constant, and equation (19) then becomes

$$\frac{1}{3}\left(2f + \eta\frac{df}{d\eta}\right) = c\frac{df}{d\eta}\left(\frac{d^2f}{d\eta^2} + \frac{1}{\eta}\frac{df}{d\eta}\right). \tag{21}$$

To solve this equation we first change the independent variable from η to z by the relation

$$z = \eta^{\frac{3}{2}}. \tag{22}$$

Then, equation (21) becomes

$$4f + 3z\frac{df}{dz} = C\left\{z\frac{df}{dz}\frac{d^2f}{dz^2} + \left(\frac{df}{dz}\right)^2\right\}, \tag{23}$$

where $C = 81c/4$.

Assuming the solution of this equation in the form of a power series of z, as

$$f = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + \dots, \tag{24}$$

where the a's are constants, and inserting this in (23), the constants have been determined, in the usual way, up to a_8. Thus, writing $\sqrt{(a_0 C)} = \alpha$ for simplicity we have

$$\frac{f}{a_0} = 1 - 2\frac{z}{\alpha} + \frac{7}{6}\left(\frac{z}{\alpha}\right)^2 - \frac{7}{216}\left(\frac{z}{\alpha}\right)^3 - \frac{77}{4320}\left(\frac{z}{\alpha}\right)^4 - \frac{1183}{103680}\left(\frac{z}{\alpha}\right)^5$$

$$- \frac{617}{77760}\left(\frac{z}{\alpha}\right)^6 - \frac{3029911}{522547200}\left(\frac{z}{\alpha}\right)^7 - \frac{20547017}{4702924800}\left(\frac{z}{\alpha}\right)^8 + \ldots \quad (25)$$

Since, however, $z = \eta^{\frac{3}{2}}$, we have

$$\frac{f}{a_0} = 1 - 2\eta_1^{\frac{3}{2}} + \frac{7}{6}\eta_1^3 - \frac{7}{216}\eta_1^{\frac{9}{2}} - \frac{77}{4320}\eta_1^6 - \frac{1183}{103680}\eta_1^{\frac{15}{2}}$$

$$- \frac{617}{77760}\eta_1^9 - \frac{3029911}{522547200}\eta_1^{\frac{21}{2}} - \frac{20547017}{4702924800}\eta_1^{12} + \quad , \quad (26)$$

where

$$\eta_1 = \frac{\eta}{\alpha^{\frac{2}{3}}} = \frac{\eta}{(a_0 C)^{\frac{1}{3}}} \quad (27)$$

With this expression for the function f, (17) gives the expression for u^*/U.

If the value of u^* at the centre of the wake be denoted by u_0^*, we readily have, since $(f)_{\eta=0} = a_0$,

$$\frac{u^*}{u_0^*} = 1 - 2\eta_1^{\frac{3}{2}} + \frac{7}{6}\eta_1^3 - \frac{7}{216}\eta_1^{\frac{9}{2}} - \frac{77}{4320}\eta_1^6 - \frac{1183}{103680}\eta_1^{\frac{15}{2}}$$

$$- \frac{617}{77760}\eta_1^9 - \frac{3029911}{522547200}\eta_1^{\frac{21}{2}} - \frac{20547017}{4702924800}\eta_1^{12} + \quad . \quad (28)$$

With the aid of this formula, the values of u^*/u_0^* have been calculated. The results are shown in Table I. The values corresponding to larger values of

TABLE I

η_1	u^*/u_0^*	θ/θ_0
0	1	1
0.1	0.938	0.953
0.2	0.830	0.868
0.3	0.703	0.761
0.4	0.568	0.639
0.5	0.437	0.508
0.6	0.318	0.370
0.7	0.219	0.257
0.8	0.146	0.117
0.9	0.099	0.031
1.0	0.075	0.002

η_1, however, have been determined by the method of numerical integration. The curve for u^*/u_0^* plotted against η_1 is shown in fig 1

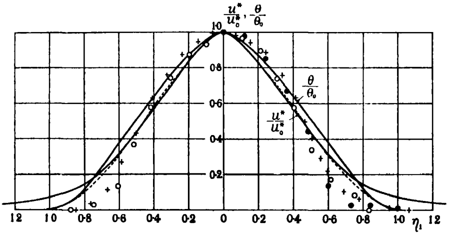


FIG 1. Velocity distribution (u^*/u_0^*) and temperature distribution (θ/θ_0) according to the modified transport theory - - - velocity distribution according to the momentum transport theory (Swain), + Schlichting's experimental points for the larger plate, ○ Schlichting's experimental points for the smaller plate, ● Simmon's experimental points

7. It will be of interest to show here that u^*/u_0^* does not vanish at any finite value of η_1.

We have, from (21)

$$\frac{d}{d\eta}(\eta^2 f) = 3c\left\{\eta\frac{df}{d\eta}\frac{d^2f}{d\eta^2} + \left(\frac{df}{d\eta}\right)^2\right\}, \tag{29}$$

and on integration

$$\eta^2 f = 3c\left\{\int_0^\eta \eta\frac{df}{d\eta}\frac{d^2f}{d\eta^2}\,d\eta + \int_0^\eta\left(\frac{df}{d\eta}\right)^2 d\eta\right\} + \text{const.} \tag{30}$$

Taking account of the boundary conditions (20), i e $(f)_{\eta=0} > 0$, $(df/d\eta)_{\eta=0} = 0$, the partial integration gives

$$\int_0^\eta \eta\frac{df}{d\eta}\frac{d^2f}{d\eta^2}\,d\eta = \tfrac{1}{2}\eta\left(\frac{df}{d\eta}\right)^2 - \tfrac{1}{2}\int_0^\eta\left(\frac{df}{d\eta}\right)^2 d\eta,$$

and therefore inserting this in (30)

$$\eta^2 f = \frac{3c}{2}\left\{\eta\left(\frac{df}{d\eta}\right)^2 + \int_0^\eta\left(\frac{df}{d\eta}\right)^2 d\eta\right\} + \text{const.} \tag{31}$$

The constant of integration can readily be determined by the conditions just referred to, and we find that it is equal to zero. Thus,

$$\eta^2 f = \frac{3c}{2}\left\{\eta\left(\frac{df}{d\eta}\right)^2 + \int_0^\eta \left(\frac{df}{d\eta}\right)^2 d\eta\right\}. \tag{32}$$

Since c is evidently positive, the right-hand side of this equation is always positive for any finite (positive) value of η, and it follows therefore that f does not vanish at any finite value of η. Thus, we find that u^*/u_0^* does not vanish at any finite value of η_1.

Further it can be shown without difficulty that the asymptotic expansion for u^*/u_0^* is of the form

$$\frac{u^*}{u_0^*} = \frac{b}{\eta_1^2} + \ldots, \tag{33}$$

where b is a constant.

CALCULATION OF TEMPERATURE DISTRIBUTION IN THE WAKE BEHIND A HEATED BODY OF REVOLUTION

8 Let θ be the difference of the temperature in the turbulent wake behind a heated body of revolution from the temperature of the fluid outside, the body being assumed, as before, to be placed in a uniform stream of velocity U parallel to the axis of the body. Then, as has been shown by Goldstein, at a large distance downstream, θ satisfies, for a first approximation, the equation

$$U\frac{\partial \theta}{\partial x} = \frac{1}{r}\frac{\partial}{\partial r}\left\{r l_r v' \frac{\partial \theta}{\partial r}\right\}. \tag{34}$$

With the assumption (12), this becomes

$$U\frac{\partial \theta}{\partial x} = -\frac{1}{r}\frac{\partial}{\partial r}\left\{r l^2 \frac{\partial u^*}{\partial r}\frac{\partial \theta}{\partial r}\right\}. \tag{35}$$

If we put

$$\theta = \frac{1}{x^{\frac{2}{3}}} F(\eta), \tag{36}$$

and make the substitutions (16), (17) and (18), we find that the function F satisfies the equation

$$\frac{1}{3}\frac{d}{d\eta}(\eta^2 F) = \frac{d}{d\eta}\left\{\eta\phi(\eta)\frac{df}{d\eta}\frac{dF}{d\eta}\right\}, \tag{37}$$

so that on integration

$$\tfrac{1}{3}\eta^2 F = \eta\phi(\eta)\frac{df}{d\eta}\frac{dF}{d\eta} + \text{const.} \tag{38}$$

But, it must be that $dF/d\eta = 0$ when $\eta = 0$, and therefore const $= 0$. Thus, we have

$$\tfrac{1}{3}\eta^2 F = \eta\phi(\eta)\frac{df}{d\eta}\frac{dF}{d\eta}. \tag{39}$$

If, as before, the mixing length is assumed to be constant over any one section of the wake, so that $\phi(\eta) = c$, we have

$$\frac{1}{F}\frac{dF}{d\eta} = \frac{\eta}{3c}\left(\frac{df}{d\eta}\right)^{-1}, \tag{40}$$

and integrating this we get

$$F = A\exp\left[\frac{1}{3c}\int_0^{\eta}\left(\frac{df}{d\eta}\right)^{-1}\eta\,d\eta\right], \tag{41}$$

where A is a constant.

Further if we put $f^* = f/a_0$ and introduce the variable η_1 by (27), we have

$$F = A\exp\left[\frac{27}{4}\int_0^{\eta_1}\left(\frac{df^*}{d\eta_1}\right)^{-1}\eta_1\,d\eta_1\right], \tag{42}$$

f^* being equivalent to u^*/u_0^*, so that its series expression is given by (28).

Thus, denoting by θ_0 the value of θ at the centre of the wake, we have by (36)

$$\frac{\theta}{\theta_0} = \exp\left[\frac{27}{4}\int_0^{\eta_1}\left(\frac{df^*}{d\eta_1}\right)^{-1}\eta_1\,d\eta_1\right]. \tag{43}$$

Remembering that $f^* = u^*/u_0^*$ and using the series (28) we get

$$\int_0^{\eta_1}\left(\frac{df^*}{d\eta_1}\right)^{-1}\eta_1\,d\eta_1 = -\left(\tfrac{2}{9}\eta_1^{\frac{3}{2}} + \tfrac{7}{54}\eta_1^3 + \tfrac{7}{72}\eta_1^{\frac{9}{2}} + \tfrac{259}{3240}\eta_1^6 + \;.\right). \tag{44}$$

This has been conveniently used for finding the values of the integral on the right-hand side of (43) for small values of η_1

For larger values of η_1, however, the values of the integral have been obtained by numerical integration by using the values of $df^*/d\eta_1$ which have been calculated by its series formula obtained from (28) by simple differentiation

Then, the values of θ/θ_0 have been obtained by (43) The results are shown in the preceding Table I and the curve for θ/θ_0 plotted against η_1 is also shown in Fig. 1.

COMPARISON WITH EXPERIMENTS

9 In his paper referred to, Goldstein has compared his theoretical curves for velocity distribution with the results of H. Schlichting's measurements

made at the Göttingen Laboratory, as well as with those of L F. G. Simmons's measurements made at the National Physical Laboratory

We shall now compare the results of the present paper with those experiments

Schlichting measured two velocity distributions in the turbulent wake behind a solid of revolution, one 100 cm. behind a circular plate of diameter d equal to 8 cm., at a Reynolds number, Ud/ν, of about $17{\cdot}4 \times 10^4$; and the other 150 cm. behind a circular plate of diameter d equal to 4 cm., with Ud/ν about $8\ 7 \times 10^4$. The plates were at right angles to the stream

The results of Schlichting's measurements are shown in fig. 1 for comparison with the theoretical curve obtained in the present paper. In this figure the theoretical curve is made to fit Schlichting's observations at $u^*/u_0^* = 0\ 5$. The experimental points for the larger plate are shown by +, while those for the smaller plate by ○. The measurements were taken along a diameter and were not quite symmetrical, and therefore the experimental points for a complete diameter have been shown. The maximum value of u^*/U was 0·0797 for the larger plate and 0·030 for the smaller one

Simmons has measured, in the N.P L. Duplex wind tunnel, the velocity distribution in the wake behind a model of the airship R 101, without fins, etc., the maximum diameter, d, of the model being 14 54 in., its overall length 80 in , and its volume 4·5 cu ft. The measurements were made in air at a wind speed of 60 ft./sec across a section of the wake 8 ft 7 in downstream of the tail of the model Simmons's measurements were along a radius, and the maximum value of u^*/U was 0 125 The experimental points are shown by ● in fig. 1 on the right-hand side, where, as before, the theoretical and experimental curves are made to coincide at $u^*/u_0^* = 0\ 5$

In fig. 1 Swain's theoretical curve for the velocity distribution is also shown for comparison by a dotted line, which is made coincide at $u^*/u_0^* = 0{\cdot}5$ with the velocity distribution curve of the present paper.

It will be seen that the agreement between the theoretical result obtained in this paper on the basis of the modified vorticity transport theory and the results of observations is not quite satisfactory, and further that there is also no satisfactory agreement with observations for Swain's result which has been obtained on the basis of the momentum transport theory

No measurements of the distribution of temperature in the turbulent wake behind a heated body of revolution are known, so that the comparison of the theoretical curve for the temperature distribution obtained in the present paper with observations must be postponed

SUMMARY

10 In the present paper, Taylor's modified vorticity transport theory of turbulence is applied to the calculation of the velocity and temperature distributions in the turbulent wake behind a solid of revolution which is placed in a uniform stream such that its axis of revolution is parallel to the direction of the undisturbed velocity

In order to carry out the calculation, it is assumed, with Prandtl, that for sufficiently high Reynolds numbers and at a sufficient distance downstream, there is geometrical and mechanical similarity in different sections of the wake and that the values of the mixing length at corresponding points in different sections are proportional to the breadths of the sections. Also, the isotropy in turbulence is assumed

Assuming the mixing length to be constant over any one section, the distribution of mean velocity is first calculated and the result is compared with the results of Schlichting's and Simmons's observations The agreement between theory and observations is not quite satisfactory.

Next, the distribution of temperature is calculated However, the comparison of the theoretical result with observations is not made, because no measurements of the distribution of temperature in the wake behind a heated body of revolution have yet been made

REFERENCES

Fage, A. 1936 *Phil. Mag* **21**, 80.
Fage, A and Townend, H. C H. 1932 *a* *Proc Roy Soc* A, **135**, 656.
—— — 1932 *b* *Rep. Memor. Aero. Res. Comm., Lond* , No. 1474
Goldstein, S. 1935 Unpublished.
Schlichting, H 1930 *Ingen.-Arch.* **1**, 533
Swain, L. M 1929 *Proc Roy Soc* A, **125**, 647
Taylor, G I. 1935 *Proc. Roy Soc* A, **151**, 494.
—— 1937 *Proc Roy. Soc.* A, **159**, 496

Application of the modified vorticity transport theory to the turbulent spreading of a jet of air

By S. Tomotika, D.Sc.

Professor of Mechanics, Osaka Imperial University, Osaka, Japan

(*Communicated by G. I. Taylor, F R S.—Received* 3 *November* 1937)

Introduction

1. The turbulent spreading of a jet of air emerging from a small circular aperture has been discussed by Tollmien (1926) on the basis of Prandtl's momentum transport theory of turbulence, by making some appropriate assumptions. He has calculated the distribution of mean velocity in the jet at distances which are sufficiently great compared with the diameter of the aperture and it has been found that the calculated distribution of mean axial velocity is in satisfactorily good agreement with observations made at Göttingen.

The same problem seems to have been discussed also by L. Howarth in an unpublished paper on the basis of Taylor's vorticity transport theory of turbulence, by assuming that not only the mean motion, but also the turbulent motion, is symmetrical about the axis of a jet. It appears, however, that the agreement between Howarth's results of calculations and the Gottingen measurements is not so good as in the case of Tollmien's results of calculations on the momentum transport theory.

In the present paper an attempt is made to apply Taylor's modified vorticity transport theory of turbulent motion (Taylor 1935, 1937) to the discussion of the turbulent spreading into the surrounding still air of a jet of air emerging from a small circular aperture. By making some simplifying assumptions, the distribution of mean velocity is calculated for any one section of the jet whose distance from the aperture is very great in comparison with the diameter of the aperture The isotropy in turbulence is also assumed It is found that the calculated distribution of mean axial velocity is in fairly good agreement with the Gottingen measurements

Equations of motion in cylindrical co-ordinates of the modified vorticity transport theory

2. Let (x, r, θ) be cylindrical co-ordinates. Then, when the mean motion is symmetrical about the x-axis so that it is independent of θ, and the fluid

is free from body force, the equations of motion in cylindrical co-ordinates according to Taylor's modified vorticity transport theory are, as shown in a previous paper (Tomotika 1937),

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial r} = -\frac{\partial}{\partial x}\left(\frac{p}{\rho}+\tfrac{1}{2}\overline{q'^2}\right)-l_r v'\left(\frac{\partial^2 v}{\partial x^2}-\frac{\partial^2 u}{\partial x\,\partial r}\right)$$

$$+l_r v'\left(\frac{\partial^2 u}{\partial r^2}-\frac{\partial^2 v}{\partial x\,\partial r}\right)+\overline{l_\theta w'}\left(\frac{1}{r}\frac{\partial u}{\partial r}-\frac{1}{r}\frac{\partial v}{\partial x}\right), \qquad (1)$$

together with a similar equation for the r-direction.

In these equations, p is the average value of the difference of the pressure from the hydrostatic pressure and ρ is the density of the fluid concerned (u, v) are the x- and r-components of velocity of the mean motion respectively, while (u', v', w') are the velocity components of the turbulent motion in the x-, r- and θ-directions and $q'^2 = u'^2+v'^2+w'^2$ Also, l_x, l_r, l_θ are the components of the mixing length l in the directions of the x-, r- and θ-axes respectively

The equation of continuity in the present case is

$$\frac{\partial}{\partial x}(ru)+\frac{\partial}{\partial r}(rv) = 0. \qquad (2)$$

TURBULENT SPREADING OF A JET OF AIR INTO THE SURROUNDING STILL AIR

3 We now proceed to the discussion of the turbulent spreading into the still air of a jet of air emerging from a small circular aperture It is assumed that the axis of the jet is perpendicular to the plane of the aperture and the mean motion in the jet is symmetrical about its axis. We consider the sections of the jet at distances which are great compared with the diameter of the aperture and we assume that the turbulence there is fully developed and isotropic so that $\overline{u'^2} = \overline{v'^2} = \overline{w'^2}$.

Then, taking the x- and r-axes along and perpendicular to the axis of the jet respectively and the origin at a certain point, the equations of motion according to the modified vorticity transport theory are given by (1), together with one similar equation.

These equations of motion are however greatly simplified if we make the usual assumptions of the boundary layer theory, namely that v and its derivatives are small in comparison with u and its corresponding derivatives;

and that derivatives with respect to x are small compared with the corresponding derivatives with respect to r. Further, in accordance with the results of the Göttingen measurements, the pressure in the jet is assumed to be constant, so that the derivatives of $(p/\rho+\frac{1}{2}\overline{q'^2})$ in the equations of motion may be neglected.

Then, for a first approximation, we have, from (1), the following equation:

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial r}=\overline{l_r v'}\frac{\partial^2 u}{\partial r^2}+\overline{l_\theta w'}\frac{1}{r}\frac{\partial u}{\partial r}. \tag{3}$$

If the mixing length l is small and the turbulence is isotropic, we have

$$\overline{l_r v'}=\overline{l_\theta w'}. \tag{4}$$

With Prandtl, we put

$$\overline{l_r v'}=\overline{l_\theta w'}=l^2\left|\frac{\partial u}{\partial r}\right|. \tag{5}$$

Then, remembering that u decreases outwards from the axis, we get

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial r}=-l^2\frac{\partial u}{\partial r}\left(\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}\right). \tag{6}$$

For the mixing length l we assume, with Tollmien, that

$$l=cx, \tag{7}$$

where c is a constant.

4. Now, the assumed constancy of the pressure in the jet requires that the momentum in the x-direction should be constant. Thus,

$$2\pi\rho\int_{-\infty}^{\infty}u^2r\,dr=\text{const} \tag{8}$$

We put

$$\eta=\frac{r}{x}, \tag{9}$$

and

$$u=\phi(x)f(\eta). \tag{10}$$

Inserting these in (8), we readily find that

$$\phi(x)=\frac{1}{x}. \tag{11}$$

Therefore

$$u=\frac{1}{x}f(\eta). \tag{12}$$

Next, the equation of continuity

$$\frac{\partial}{\partial x}(ru)+\frac{\partial}{\partial r}(rv)=0 \tag{13}$$

is satisfied by

$$ru=-\frac{\partial\psi}{\partial r},\quad rv=\frac{\partial\psi}{\partial x}, \tag{14}$$

where ψ is Stokes's current function.

The expression for ψ can easily be obtained by combining (12) with the first equation in (14). We have

$$\psi=-x\int\eta f(\eta)\,d\eta. \tag{15}$$

Thus, writing

$$\int\eta f(\eta)\,d\eta=F(\eta), \tag{16}$$

we have

$$\psi=-xF(\eta). \tag{17}$$

Putting this in (14) and taking (9) into account, we readily have

$$\left.\begin{aligned} u&=\frac{1}{x\eta}\frac{dF}{d\eta},\\ v&=\frac{1}{x}\frac{dF}{d\eta}-\frac{F}{x\eta}.\end{aligned}\right\} \tag{18}$$

The substitution of these in (6) gives immediately a differential equation for the function $F(\eta)$. With the assumption (7) for the mixing length we have

$$F\frac{d^2F}{d\eta^2}+\left(\frac{dF}{d\eta}\right)^2-\frac{1}{\eta}F\frac{dF}{d\eta}=c^2\left(\frac{d^2F}{d\eta^2}-\frac{1}{\eta}\frac{dF}{d\eta}\right)\left(\frac{d^3F}{d\eta^3}-\frac{1}{\eta}\frac{d^2F}{d\eta^2}+\frac{1}{\eta^2}\frac{dF}{d\eta}\right). \tag{19}$$

Now, $\eta=0$ corresponds evidently with $r=0$, and the conditions for u and v at $\eta=0$ are evidently

$$(u)_{\eta\to0}\neq0,\quad (v)_{\eta\to0}=0. \tag{20}$$

Hence, for small values of η, the function F must be such that

$$F\propto\eta^2, \tag{21}$$

as can easily be seen from the formulae (18) for (u,v)

Thus, we may write

$$F=\eta^2F_*, \tag{22}$$

the condition at $\eta=0$ for the function F_* being now

$$(F_*)_{\eta\to0}\neq0. \tag{23}$$

We put (22) in (19). Then, the equation for F_* becomes

$$4F_*^2 + 7\eta F_* \frac{dF_*}{d\eta} + \eta^2\left(\frac{dF_*}{d\eta}\right)^2 + \eta^2 F_* \frac{d^2F_*}{d\eta^2} = \frac{c^2}{\eta}\left(3\frac{dF_*}{d\eta} + \eta\frac{d^2F_*}{d\eta^2}\right)\left(3\frac{dF_*}{d\eta} + 5\eta\frac{d^2F_*}{d\eta^2} + \eta^2\frac{d^3F_*}{d\eta^3}\right) \tag{24}$$

A little calculation shows that the function F_* may be expanded in a power series of $\eta^{\frac{3}{2}}$. Thus, putting

$$z = \eta^{\frac{3}{2}}, \tag{25}$$

and using this relation, we first change the independent variable from η to z in the differential equation (24). We then have

$$F_*^2 + \tfrac{45}{16} z F_* \frac{dF_*}{dz} + \tfrac{9}{16} z^2\left(\frac{dF_*}{dz}\right)^2 + \tfrac{9}{16} z^2 F_* \frac{d^2F_*}{dz^2} = \tfrac{27}{128} c^2\left(7\frac{dF_*}{dz} + 3z\frac{d^2F_*}{dz^2}\right)\left(7\frac{dF_*}{dz} + 13z\frac{d^2F_*}{dz^2} + 3z^2\frac{d^3F_*}{dz^3}\right) \tag{26}$$

5 Next, assuming the solution of this equation in the form of a power series of z, as

$$F_* = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 + \dots, \tag{27}$$

where the a's are constants, and inserting this in (26), the constants have been determined, in the usual manner, up to a_5.

Thus, we have

$$\frac{F_*}{a_0} = 1 - \frac{8}{21}\frac{\sqrt{2}\,z}{\sqrt{3}\,c} + \frac{22}{405}\left(\frac{z}{c}\right)^2 - \frac{1363}{199017}\frac{\sqrt{2}}{\sqrt{3}}\left(\frac{z}{c}\right)^3 + \frac{592111}{1432922400}\left(\frac{z}{c}\right)^4 - \frac{40780921}{1429340094000}\frac{\sqrt{2}}{\sqrt{3}}\left(\frac{z}{c}\right)^5 + \dots \tag{28}$$

Since, however, $z = \eta^{\frac{3}{2}}$, we get

$$\frac{F_*}{a_0} = 1 - \frac{8}{21}\frac{\sqrt{2}}{\sqrt{3}}\left(\frac{\eta}{\alpha}\right)^{\frac{3}{2}} + \frac{22}{405}\left(\frac{\eta}{\alpha}\right)^3 - \frac{1363}{199017}\frac{\sqrt{2}}{\sqrt{3}}\left(\frac{\eta}{\alpha}\right)^{\frac{9}{2}} + \frac{592111}{1432922400}\left(\frac{\eta}{\alpha}\right)^6 - \frac{40780921}{1429340094000}\frac{\sqrt{2}}{\sqrt{3}}\left(\frac{\eta}{\alpha}\right)^{\frac{15}{2}} + \dots, \tag{29}$$

with $\alpha = c^{\frac{2}{3}}$.

Using this, together with (18) and (22), the velocity components u and v can be calculated Since, however, we are mainly interested in the distribution of the axial component of velocity, u, measurements of which have been made at Göttingen, only the expression for u will be obtained here.

By (18) and (22), we have

$$\frac{ux}{2a_0} = \frac{F_*}{a_0} + \frac{\eta}{2a_0}\frac{dF_*}{d\eta}, \tag{30}$$

and substituting (29) in this, we get

$$\frac{ux}{2a_0} = 1 - \frac{2\sqrt{2}}{3\sqrt{3}}\eta_1^{\frac{3}{2}} + \frac{11}{81}\eta_1^3 - \frac{1363}{61236}\frac{\sqrt{2}}{\sqrt{3}}\eta_1^{\frac{9}{2}} + \frac{592111}{358230600}\eta_1^6 - \frac{40780921}{300913704000}\frac{\sqrt{2}}{\sqrt{3}}\eta_1^{\frac{15}{2}} + \dots, \tag{31}$$

where we have put for simplicity

$$\eta_1 = \frac{\eta}{\alpha} = \frac{\eta}{c^{\frac{1}{3}}}. \tag{32}$$

If the value of u on the axis of the jet be denoted by u_0, we readily have, since $u_0 x = (ux)_{\eta=0} = 2a_0$,

$$\frac{u}{u_0} = 1 - \frac{2\sqrt{2}}{3\sqrt{3}}\eta_1^{\frac{3}{2}} + \frac{11}{81}\eta_1^3 - \frac{1363}{61236}\frac{\sqrt{2}}{\sqrt{3}}\eta_1^{\frac{9}{2}} + \frac{592111}{358230600}\eta_1^6 - \frac{40780921}{300913704000}\frac{\sqrt{2}}{\sqrt{3}}\eta_1^{\frac{15}{2}} + \dots \tag{33}$$

This series has been conveniently used for the calculation of the values of u/u_0 for small values of η_1, and the calculation has been continued, for larger values of η_1, by numerical solution of the differential equation. The curve for u/u_0 plotted against η_1 is shown in fig. 1.

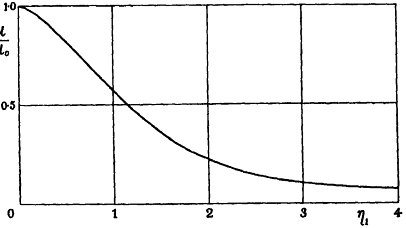


FIG. 1. Velocity distribution according to the modified vorticity transport theory.

6. It will be of interest to show here that u/u_0 does not vanish at any finite value of η_1.

Now, the equation (19) can be expressed in the form

$$\frac{d}{d\eta}\left(\frac{1}{\eta}F\frac{dF}{d\eta}\right)=\frac{c^2}{2\eta}\frac{d}{d\eta}\left\{\left(\frac{d^2F}{d\eta^2}-\frac{1}{\eta}\frac{dF}{d\eta}\right)^2\right\} \tag{34}$$

Integrating this once, we readily get

$$\frac{1}{\eta}F\frac{dF}{d\eta}=\frac{c^2}{2}\left\{\frac{1}{\eta}\left(\frac{d^2F}{d\eta^2}-\frac{1}{\eta}\frac{dF}{d\eta}\right)^2+\int_0^\eta\frac{1}{\eta}\left(\frac{d^2F}{d\eta^2}-\frac{1}{\eta}\frac{dF}{d\eta}\right)^2d\eta\right\}. \tag{35}$$

Evidently the right-hand side of this equation does not vanish at any finite (positive) value of η, and therefore $dF/d\eta$ does not vanish at any finite value of η, except at $\eta = 0$. Thus, with the aid of (18), we find that u/u_0 does not vanish at any finite value of η_1.

Further it can be shown without difficulty that the asymptotic expansion for u/u_0 is of the form

$$\frac{u}{u_0}=\frac{b}{\eta_1}+ \tag{36}$$

COMPARISON WITH OBSERVATION

7. The distribution of mean axial velocity in a jet of air emerging from a small circular aperture has been measured at the Gottingen Laboratory, at distances which are great compared with the diameter of the aperture. In his paper already referred to, Tollmien has compared the result of his calculation on the basis of the momentum transport theory with the results of the Göttingen measurements and it has been found that the agreement between theory and observations is fairly satisfactory

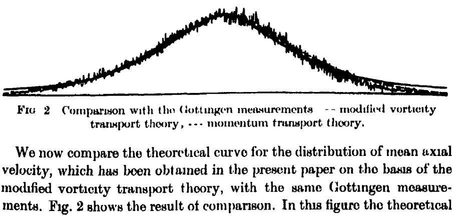


FIG 2 Comparison with the Gottingen measurements -- modified vorticity transport theory, --- momentum transport theory.

We now compare the theoretical curve for the distribution of mean axial velocity, which has been obtained in the present paper on the basis of the modified vorticity transport theory, with the same Gottingen measurements. Fig. 2 shows the result of comparison. In this figure the theoretical curve is made to coincide with the experimental curve at $u/u_0 = 0{\cdot}5$. It will be seen that the agreement between theory and observations is fairly satisfactory, as in Tollmien's case.

In conclusion the writer wishes to express his thanks to Mr I Imai for his assistance in the calculations of the present paper.

SUMMARY

8 The turbulent spreading of a jet of air emerging from a small circular aperture is discussed on the basis of Taylor's modified vorticity transport theory of turbulent motion Assuming the isotropy in turbulence, the distribution of mean axial velocity is calculated for any one section of the jet whose distance from the aperture is great in comparison with the diameter of the aperture. The calculated curve is compared with the Göttingen measurements and the fairly satisfactory agreement between theory and observations is found, as in the case of Tollmien's calculation on the basis of Prandtl's momentum transport theory.

REFERENCES

Taylor, G. I. 1935 *Proc. Roy Soc* A, **151**, 494
—— 1937 *Proc Roy Soc* A, **159**, 496.
Tollmien, W. 1926 *Z. angew. Math Mech* **6**, 468.
Tomotika, S. 1937 *Proc. Roy. Soc* A, **165**, 53

An experimental determination of the spectrum of turbulence

By L. F. G. Simmons, M.A. and C. Salter, M.A.

With an Appendix. Method of Deducing $F(n)$ from the Measurements

By G. I. Taylor, F.R.S.

The experimental investigation of isotropic turbulent motion is most conveniently conducted in the air stream of a wind tunnel in which the turbulence is augmented by the addition of a grid of uniform mesh placed across it. At any point downstream, beyond the wind shadow of the grid, the velocity fluctuations are small compared with the mean speed of the stream, but vary irregularly with time. Records taken of u, the instantaneous value of the turbulent component in the direction of motion, show that the time variations follow the random law of errors (Simmons and Salter 1934, Townend 1934), but no successful attempts appear to have been made to analyse such a record into its harmonic components.* On the other hand, by utilizing a hot-wire anemometer to produce a current proportional to u, and employing electrical filters to measure the contributions to $\overline{u^2}$ which arise from frequencies within the range 0 to n where n is varied, data may be obtained from which a spectrum curve of the variation in velocity u, at a fixed point can be plotted. The ordinate of this curve at frequency n represents $F(n)$, the function denoting the probability of the existence of velocities between n and $n+dn$, whose significance is discussed by Professor G. I. Taylor in the paper which follows.

In some experiments undertaken at Professor Taylor's suggestion for the purpose of determining the values of $F(n)$ for turbulence created by a grid of square mesh, the hot-wire technique was employed to measure $\overline{u^2}$ from the readings of a thermal milliammeter, placed in the output circuit of a valve amplifier. In order to effect the analysis different electrical filters were successively put in the measuring circuit. Two types of filter were used. one, a *low-pass*, allowed the passage only of currents below a certain frequency, the other, a *high-pass*, only those above a certain critical frequency, the limit in each case being settled by the electrical constants of the filter. The former type served to explore the range 0–325 c./sec.; the latter was used for frequencies above this figure From the ratios of readings of the milliammeter taken, in each case with and without the filter, it was possible to obtain, at any given wind speed, a close approximation to the spectrum curve required.

---

* While this work was in progress an attempt in this direction was made jointly by Dryden, Schaubauer, Mock and Skramstad, see Nat. Adv. Cttee. Aero., *Report* No. 581 (1937).

EXPERIMENTAL DETAILS

The Wind Tunnel

The 4 ft. N P.L wind tunnel used for the present work has a bell-mouthed inlet fitted with a honeycomb for the purpose of guiding the air and at the same time preventing any axial rotation of the stream Over the greater part of the working section the velocity distribution is fairly uniform to within 3 or 4 in from the walls, the general steadiness is good, but influenced by draughts in the room Experience shows that the most favourable conditions for studying turbulent motion are provided when a grid of uniform mesh, opposing a high resistance to the flow, is placed across the tunnel, for the turbulence downstream is then less susceptible to the influence of outside disturbances and at the same time is more uniformly distributed across a transverse section. An examination of the flow immediately behind a square-mesh grid of thin slats reveals a regular system of eddies rotating about horizontal and vertical axes. During the course of their passage downstream the eddies mix with the high velocity jets issuing through the mesh There is some evidence to show that the mixing process continues for a distance of about 15 times the space length between the centres of the slats, after which the stream is statistically uniform, the individual wakes having disappeared, and the mean velocity at all points in a cross-section becomes sensibly constant Moreover, the turbulence is isotropic, and has a scale, as Professor Taylor* has shown, determined by the mesh-length of the grid

In earlier experiments undertaken to verify some of the predicted properties of this particular type of turbulence, measurements were made in the wake of a grid composed of thin, metal slats 1 in. wide attached to the upstream side of a honeycomb of square cells of 3 in. mesh, similar in all respects to that fixed in the inlet of the tunnel. With the same grid used for creating turbulence, the present enquiry was confined to a study of the velocity variations at a point on the axis of the tunnel 6·83 ft. from the grid, where the intensity of turbulence, expressed as the ratio $\sqrt{\overline{u^2}}/U$, was 0·0296

---

* *J. Aero. Sci.* No. 8, 4 (1937).

The Electrical Circuits

Since the principal features of the electrical circuits have been previously described in a paper on the measurements of u^2 for air-flow through a pipe,* it will suffice here to give a short outline of the general arrangement.

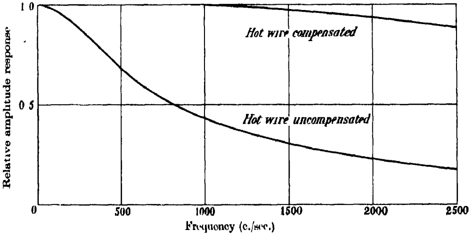
First, mention should be made of the Wheatstone bridge in which the wire is heated and maintained on the average about 200° above the temperature of the stream After the current is adjusted to bring the bridge into balance, as indicated by a long-period galvanometer, turbulence changes the temperature, and hence the resistance, of the wire The accompanying potential variations across the wire produce fluctuating out-of-balance potentials, which at any instant would be proportional to the turbulent velocity component in the direction of motion were it not for a lag in the temperature response of the wire at high frequencies For the small wires of 0·0001 in diameter used in the experiments, the effect of the lag is negligible up to a frequency of 100 per sec Within this range there is therefore an almost exact correspondence between the potential variations and the fluctuations of speeds But with rise of frequency the lag produces a loss of amplitude and change of phase, both of which increase progressively in magnitude. A close approach to the ideal condition of uniform response can, however, be obtained over a fairly wide range of frequencies, by a method described by Dryden and Kuethe (1929), in which the out-of-balance potential is applied to a valve amplifier having a compensating circuit in one of the stages. The circuit consists of a resistance with an inductance in series, the potential across it is passed on to the next stage, so that the change of impedance gives increased amplification and change in phase with frequency which serve to neutralize the losses incurred through thermal lag. Thus, provided all the component frequencies are within the effective range of the compensation applied, the output current from the amplifier will indicate in detail the time variation of the turbulent velocity

An indication of the effectiveness of the compensation applied to one of the wires used at a mean wind speed of 20 ft./sec , is afforded by a comparison of the curves of fig 1 These show the relative amplitude of response at different frequencies due to a given sinusoidal variation of velocity, the lower curve refers to the uncompensated wire, the upper curve to the wire with compensation introduced, and thus includes the losses due to the distortion of the amplifier.

The audio-frequency amplifier used for the present work was the one described in the paper already referred to. It consists of three stages, with resistance-capacity coupling. Of these the first and third are amplifying stages, the second being reserved for the compensating circuit, connected between the plate of the second valve and the H.T. supply. In its earlier form, with coupling condensers of $0\cdot1\,\mu$F and 1 MΩ grid leaks, the overall amplification—determined from measurements of the mean-square output current produced by a known voltage applied to the input—was uniform from about 20 to 1000 c./sec. The lack of response at the lower end acts advantageously in discriminating between the slow variations of mean speed arising from the irregular running of the fan, and the faster variations associated with turbulence generated by the grid, and also present in the disturbances entering the tunnel and passing freely through the mesh.

---

* *Rep. Memor. aero. Res. Comm., Lond* , No. 1651 (1934).

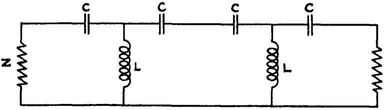


Since preliminary measurements made with different coupling condensers showed that an appreciable amount of turbulent energy existed in the low frequency end of the spectrum, in order to increase the amplification in this region the amplifier had the same coupling condensers as before, but larger grid leaks of 2 MΩ. By such means errors arising from the non-uniform response of the wire and amplifier combination, which lead to an underestimation of $\overline{u^2}$, were, under the conditions prevailing at the lowest wind speed, reduced from 5 to 1·5 %.

To eliminate the risk of back coupling which sometimes occurred when a single battery was used as the common source of H.T. supply, the plate circuits of the first three stages were separated from the last and connected to another H.T. battery. An additional change involved the re-arrangement of the output circuit, to allow of the inclusion of the electrical filters used in the course of the analysis.

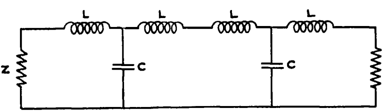
The Electrical Filters

These were of two kinds one a low-pass, capable of transmitting currents of all frequencies below a given maximum; the other a high-pass, which allows the passage of frequencies only above a certain minimum value Each filter was made up of two sections of similar T-shaped networks connected in cascade. Fig 2 illustrates the arrangement of the sections,



Two stage low-pass filter

FIG. 2



Two stage high-pass Filter

FIG. 3

and the constituent elements, comprising two inductances, L, and a condenser, C, of a typical low-pass filter Such a filter when terminated at each end by an impedance, z, may be considered as part of an infinite chain of T sections. If no energy were dissipated in any section, on an alternating voltage being applied to the input side, currents below the cut-off frequency, $\frac{\omega_c}{2\pi}$, where $\omega_c = \sqrt{2/LC}$, are transmitted without attenuation. In practice, however, it is found that this ideal condition cannot be realized owing to the losses arising from the resistances of the various components.

As a first stage in the design of a filter to cut off at a given frequency, L and C were chosen to satisfy the equation above. After the characteristic impedance z had been calculated from the formula $z = \sqrt{\frac{2L}{C}\left(1 - \frac{1}{4\omega_c^2}\right)}$ the networks were assembled and terminated at each end by a resistance equal to z The response curve of the filter was then determined from measurements of the mean-square output current corresponding to a known voltage input, applied at different frequencies * Although the curve usually indicated an approximately constant response over a reasonable frequency range, for some of the filters (by means of a smaller resistance across the output end) an increased range followed by a steeper fall of the curve in the neighbourhood of the cut-off frequency could be obtained, the most suitable resistance for every filter was found by trial.

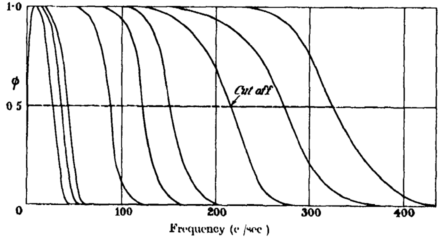


FIG. 4 Characteristic curves of low-pass filters.

Characteristic curves for most of the low-pass filters are plotted in Fig. 4, the ordinate ϕ at any frequency n being proportional to the mean-square values of the output currents for a given input voltage In its application to the present problem the filter is assumed to have the same performance as an ideal filter cutting off at a frequency where $\phi = 0{\cdot}5$, it is therefore assumed to transmit uniformly currents of lower frequency while stopping all above On interpreting the results obtained with such a filter, due allowance can be made, if necessary, for the true shape of the characteristic

---

* The beat-tone oscillator used to provide the small alternating voltage, consisted of two oscillatory circuits, one having a fixed frequency of 100,000 c /sec. and a second whose frequency could be varied, at will, between 100,000 and 110,000 c /sec By adjusting the constants of the second circuit, it was possible to obtain any beat frequency between 0 and 10,000 c /sec. The voltage actually applied across the input impedance z of the filter was obtained by amplifying the change of potential generated across a fixed resistance in the output circuit of the oscillator, and was therefore proportional to the current passing through it.

curve, but as will be shown later, the corrections involved are small, and lie well within the limits of observational error.

High-pass filters employed in the analysis of the upper part of the spectrum each consisted of two sections (see fig. 3) closed by terminal resistances. The cut-off frequency for non-dissipative networks of this type is given by the relation $\omega_c = \sqrt{2/LC}$, whilst the theoretical value of the impedance for currents of frequency $\frac{\omega}{2\pi}$ is $z = \sqrt{\frac{2L}{C}\left\{1-\left(\frac{\omega_c}{\omega}\right)^2\right\}}$ Here again the formulae were useful as a guide, although, as in the case of the low-pass filters, the most suitable values for the output impedances had to be found experimentally Fig 5 shows the measured response curves for the high-pass filters each of which was assumed to be equivalent to an ideal filter having a cut-off frequency at $\phi = 0.5$.

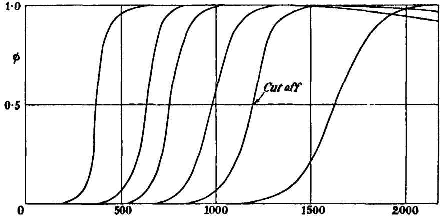


Fig. 5. Characteristic curves of high-pass filters

Measurements of the Spectrum of Turbulence

(a) *Method* The method of using a filter to determine the proportion of $\overline{u^2}$ contributed by the band of frequencies dealt with must now be described. Suppose the wire held in the stream and the heating current adjusted to bring the Wheatstone bridge into balance at the mean speed of flow. Speed variations due to turbulence produce out-of-balance potentials: on being amplified these give rise to the fluctuating currents in the anode circuit of the last valve The filter is connected across a resistance in the anode circuit and the mean-square value of the output current $\overline{i_1^2}$ is measured. Writing $\overline{u_1^2}$ for the mean of the square of the turbulent components contained within the range from 0 to N_1, the cut-off frequence of the filter,* we have

$$\overline{i_1^2} = K_1 \overline{u_1^2},$$

where K_1 is the product of the amplification factor and the calibration constant of the wire Next, suppose the filter replaced by a non-inductive resistance Let i_2^2 be the observed mean-square value of the current, then, since the resistance possesses no selective properties and passes all frequencies uniformly,

$$\overline{i_2^2} = K_2 \overline{u^2},$$

where K_2 is defined as above Provided the two sets of measurements are taken within a short time of each other, that is, while the amplification is constant and before the calibration of the wire is changed by dust accidentally adhering to it,† with the resistance suitably chosen K_1 may be made equal to K_2. Hence, from the equations above

$$\frac{\overline{u_1^2}}{\overline{u^2}} = \frac{\overline{i_1^2}}{\overline{i_2^2}}$$

It will be apparent therefore that, since the ratio of two current readings provide a measure of the fractional part of turbulent energy distributed within a given band of frequencies, namely from 0 to N_1, similar observations taken with a number of filters, each having a different cut-off frequency, will furnish the information needed for constructing a spectrum curve.

The arrangement of the output circuit of the amplifier, including for purposes of illustration a low-pass filter, is shown in fig. 6 Other details represented include the input resistance R_2, equal to the calculated value of z for the filter, the output terminal resistance R_3, and a resistance R_1 provided as a substitute for the filter during the measurements of $\overline{u^2}$. The method of connecting, alternatively, the filter and the circuit $R_1 R_3$ across R_2, will be clear from an inspection of the diagram, as will also the method of balancing the steady drop of potential across R_2 by means of a circuit comprising a battery E and galvanometer G_2. Both $\overline{i_1^2}$ and $\overline{i_2^2}$ were indicated by the deflection of the mirror galvanometer connected to the thermo-junction element, T J, of an A C milliammeter. Except when high-pass filters were employed, the current generally varied slowly but irregularly within limits of ± 4 %, in unison with the casual disturbances present in the stream. Therefore, in order that the results should afford a measure of the average turbulence, the mean deflection in each case was estimated from observations extending over a period from 3 to 5 min. The current through R_1R_3 resulting from the application of a known sinusoidal voltage to the input side, served to check the performance of the amplifier.

---

* In accordance with the assumption previously made, $\phi = 1$ from $n = 0$ to N_1, and zero from $n = N_1$ to ∞.

† Professor Taylor has pointed out that the conditions in the return flow tunnel at Cambridge are more favourable in this respect and do not lead to a continuous change in the calibration, such as we found to occur with every wire used in the present experiments.

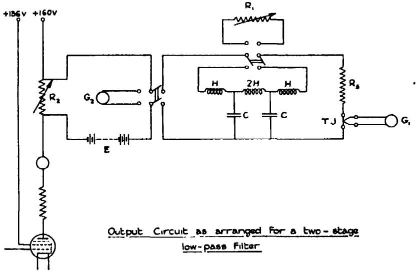


FIG. 6

The spectrum measurements were made at mean wind speeds of 15, 20, 25, 30 and 35 ft /sec, low-pass filters being used to explore the range below 325 c /sec. Apart from effects due to accidental changes in the calibration of the wire, the results obtained were generally in error by less than ± 4 %. But at higher frequencies, because of improved accuracy, high-pass filters were used to measure $1-\overline{u_1^2}/\overline{u^2}$, the errors thereby being reduced to within the limits of ± 1 %.

(*b*) ***Results***. The values of $\overline{u_1^2}/\overline{u^2}$ measured in the experiments at the various wind speeds together with smooth curves drawn to lie evenly between the observations are shown in figs. 7 and 8. In view of the assumptions made in deducing the results, it will be necessary first to consider the corrections required when account is taken of the true form of the

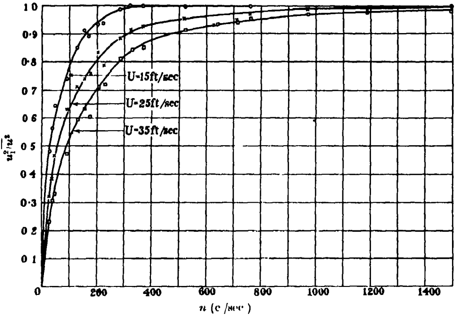

1 0
0·9
0·8
0 7
0 6
0 5
0 4
0·3
0 2
0 1
0
200
400
600
800
1000
1200
1400
u_1^2/u^2
n (c /sec)
U=15ft/sec
U=25ft/sec
U=35ft/sec

FIG 7

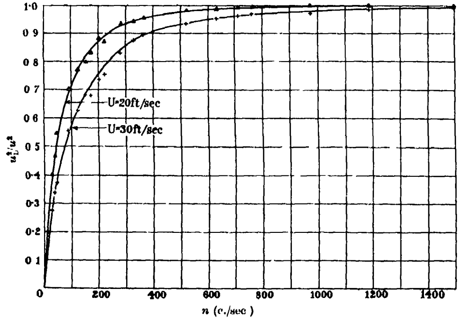

1·0
0 9
0 8
0 7
0 6
0 5
0·4
0·3
0·2
0 1
0
200
400
600
800
1000
1200
1400
u_1^2/u^2
n (c./sec)
U=20ft/sec
U=30ft/sec

FIG 8

characteristic function for each filter, and for the loss of response of the wire-amplifier combination, especially noticeable in the neighbourhood $n = 1$. A method of estimating the correction ϵ_1 is described in the Appendix where it is shown that

$$\epsilon_1 = \frac{\overline{u_1^2}}{\overline{u^2}} - \int_0^\infty F_1(n)\,\phi_1(n)\,dn$$

The function $F_1(n)$ is the first approximation to the spectrum function $F(n)$ and is obtained by graphical differentiation of the slope of the smooth curve, representing the observations at any given wind speed. Although $F_1(n)$ can readily be found where the curve is well defined, owing to the absence of observations below $n = 26$, its value in this region is subject to some uncertainty.

In the application of the method to the results at $U = 20$ ft./sec. the form of the correction was slightly modified to allow for the fact that all measurements of $\overline{u_1^2}$ and $\overline{u^2}$ were virtually made with low-pass filters cutting off at the lower end at $n = 0{\cdot}8$. Since such filters are unable to detect turbulence for which $n < 0{\cdot}8$, the formula must only be applied from $n = 0{\cdot}8$ to ∞. Instead of ϵ_1 Professor Taylor proposed a new correction

$$\epsilon_1' = \frac{\overline{u_1^2}}{\overline{u^2}} - \int_{0{\cdot}8}^{\infty} F_1(n)\,\phi_1(n)\,dn.$$

This was calculated at a number of points from the values of $F_1(n)$ taken from the curve extended below $n = 26$ to cut the axis at 0·8. The results are tabulated below together with the observed and smoothed values of $\overline{u_1^2}/\overline{u^2}$.

Cut-off frequency c./sec.	Type of filter	Values of $\overline{u_1^2}/\overline{u^2}$		
		Observed	Smoothed	ϵ_1
26	Low pass	0·40	0·40	0·005
34·5		0·46	0·46	0·007
43		0·55	0·51	0·009
87		0·69	0·68	0·002
122		0·77	0·78	−0·003
151		0·80	0·82	−0·002
275		0·93	0·93	0·001
325		0·94	0·95	−0·003
365	High pass	0·955	0·954	0
630		0·987	0·986	0·001
760		0·993	0·993	0·001
980		0·997	0·997	0
1190		0·999	0·999	0·001

Except at the lowest frequencies the corrections are seen to be negligible in magnitude and variable in sign. Mainly for these reasons and because of the uncertainty arising from the unavoidable lack of observations below $n = 26$, the corrections are not included in the results appended.

The curves in figs. 6 and 7 and the figures in Table I are therefore taken to represent the statistical distribution of the intensity of isotropic turbulence created by the grid One feature disclosed by the measurements is the high proportion of energy existing in the lower frequencies, for example, the contribution due to components in the range 0–100 c./sec. amounts to 0·8 at $U = 15$, and to 0 53 at $U = 35$ ft /sec. Another feature relates to the frequencies present, for, whereas at the lowest speed the highest recorded frequency was in the neighbourhood of 600, at the highest speed it exceeded 2500 c /sec Chief interest, however, centres in the statistical frequency function $F(n)$ derived from the curves. This function is tabulated for the different speeds in Table II, and also exhibited graphically by the curves in fig. 9

In conclusion, the writers desire to express their indebtness to Professor Taylor for his helpful suggestions and advice, and to acknowledge Mr W. G. Raymer's assistance in making some of the observations

SUMMARY

The time variation of velocity at a fixed point in a turbulent air stream is analysed into a spectrum The method adopted involves the use of the ordinary hot-wire technique to produce changes of potential in a Wheatstone bridge circuit, which are magnified by a valve amplifier The fluctuating voltage drop generated across a resistance in the output circuit of the amplifier is then applied, in turn, to electrical filters having different cut-off frequencies In each case the output current is measured, with and without each filter in circuit, by means of a thermal milliammeter which indicates the mean value of the square of the current supplied to it From the ratios of the readings taken with and without each filter, the spectrum curve is calculated by a method described in the Appendix.

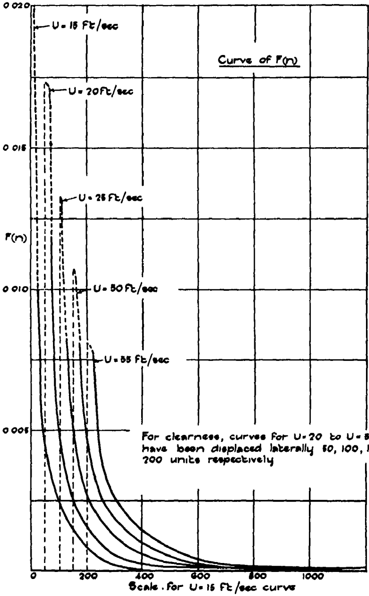
All measurements were made in a wind tunnel, at a point in the air stream where the turbulence created by a grid of regular mesh was known to be isotropic the wind speeds used were 15, 20, 25, 30 and 35 ft./sec

TABLE I

n c./sec.	$\overline{u_1^2}/u^2$ $U = 15$ ft /sec.	$U = 20$ ft /sec.	$U = 25$ ft /sec.	$U = 30$ ft /sec	$U = 35$ ft /sec.
20	0 42	0 33	0 26	0 20	0 16
40	0 587	0 49	0 42	0 37	0 31
60	0 67	0 59	0 52	0 52	0 40
80	0 74	0 66	0 59	0 52	0 47
100	0 80	0 72	0 64	0 58	0 53
120	0 84	0 76	0 69	0 63	0 57
140	0 88	0 80	0 73	0 67	0 62
160	0 91	0 83	0 76	0 70	0 65
180	0 93	0 85	0 78	0 73	0 68
200	0 95	0 87	0 81	0 76	0 71
220	0 96	0 89	0 83	0 78	0 74
240	0 98	0 91	0 85	0 80	0 76
260	0 98	0 92	0 87	0 82	0 79
280	0 99	0 93	0 88	0 84	0 81
300	0 99	0 936	0 898	0 85	0 82
320	0 993	0 944	0 909	0 87	0 84
340	0 995	0 948	0 918	0 880	0 85
360	0 996	0 953	0 925	0 891	0 86
380	0 997	0 957	0 93	0 899	0 87
400	0 997	0 960	0 936	0 906	0 880
450	0 997	0 968	0 945	0 921	0 893
500	0 997	0 974	0 95	0 931	0 905
550	0 998	0 979	0 96	0 940	0 918
600	0 998	0 984	0 966	0 948	0 927
650	0 999	0 988	0 972	0 954	0 937
700	1 00	0 990	0 976	0 961	0 944
750	—	0 992	0 981	0 967	0 953
800		0 994	0 985	0 971	0 958
850	—	0 995	0 987	0 975	0 962
900		0 996	0 989	0 978	0 965
950		0 996	0 991	0 981	0 969
1000	—	0 997	0 992	0 983	0 972
1100		0 998	0 994	0 987	0 977
1200	—	0 999	0 995	0 989	0 981
1300		0 999	0 996	0 991	0 985
1400	—	1 000	0 997	0 992	0 988
1500	—	—	0 998	0 994	0 990
1600	—	—	0 998	0 995	0 992
1700		—	0 999	0 996	0 993
1800		—	0 999	0 997	0 995
1900	—	—	0 999	0 997	0 996
2000		—	1 000	0 998	0 996
2100	—	—		0 999	0 997
2200	—	—	—	0 999	0 998
2300	—	—		0 999	0 998
2400	—	—		1 000	0 998
2500	—	—	—	—	1 000

REFERENCES

Dryden and Kuethe 1929 Nat Adv Cttee. Aero, *Report* No. 320.
Simmons and Salter 1934 *Proc Roy. Soc.* A, **145**, 212.
Townend 1934 *Proc Roy Soc.* A, **145**, 180.



n (c /sec)

FIG. 9

TABLE II

n c./sec.	$F(n)$ $U = 15$ ft./sec.	$U = 20$ ft./sec.	$U = 25$ ft./sec.	$U = 30$ ft./sec.	$U = 35$ ft./sec.
5	0·0226	0·0173	0·0135	0·0103	0·010
25	0·0101	0·0093	0·0083	0·0091	0·0076
45	0·0047	0·0051	0·0056	0·0052	0·0052
65	0·0037	0·0039	0·0037	0·0034	0·0036
85	0·0028	0·0030	0·0030	0·0030	0·0031
105	0·0023	0·0024	0·0024	0·0024	0·0025
125	0·0020	0·0020	0·0020	0·0020	0·0021
145	0·0017	0·0016	0·0017	0·0018	0·0020
165	0·0013	0·0013	0·0014	0·0016	0·0014
185	0·0009	0·0011	0·0013	0·0014	0·0014
205	0·0008	0·0009	0·0012	0·0013	0·0012
225	0·0006	0·0007	0·0011	0·0010	0·0013
245	0·0004	0·0007	0·0009	0·0011	0·0012
265	0·0003	0·0006	0·0008	0·0009	0·0011
285	0·0002	0·0004	0·0007	0·0007	0·0010
310	0·0001	0·0004	0·0005	0·0007	0·00085
330	0·000075	0·0002	0·00045	0·0006	0·00075
350	0·000065	0·00025	0·00035	0·00055	0·00055
370	0·000035	0·00020	0·00030	0·00040	0·00040
390	0·000020	0·00017	0·00025	0·00035	0·00035
425	0·00001	0·00016	0·00020	0·0003	0·000275
475	0·00001	0·00012	0·00015	0·0002	0·00024
575	0·00001	0·000082	0·00012	0·00016	0·00018
675		0·000048	0·000088	0·00014	0·00014
775		0·000030	0·000080	0·000086	0·00009
875		0·000018	0·000036	0·000066	0·000074
975		0·000014	0·000026	0·000044	0·000058
1075		0·000010	0·000018	0·000028	0·000048
1275	-	0·000004	0·000010	0·000018	0·000034
1475		0·0000011	0·000006	0·000012	0·000022
1675	-	0·0000002	0·000004	0·000010	0·000014
1875	--	--	0·000004	0·000006	0·000010
2075		-	0·000002	0·000006	0·000006
2475	—			—	0·000004

APPENDIX

METHOD OF DEDUCING $F(n)$ FROM THE MEASUREMENTS

In the measurements described by Mr Simmons a fluctuating current which is proportional to the fluctuations of velocity at a fixed point in an air stream is modified by the action of a filter which passes only certain ranges of frequency.

If an alternating current of constant amplitude but variable frequency is passed through the filter the amplitude of the output will vary with the frequency. If $\phi(n)$ represents the ratio of the square of the amplitude of the output current to that of the input when the frequency is n c./sec., then $\phi(n)$ is a characteristic function for the filter. $\phi(n)$ is determined by measuring the output when an alternating current of known amplitude and frequency is applied to it.

When fluctuating currents are applied the output can only be calculated if both $\phi(n)$ and the spectrum function $F(n)$ is known If $\overline{u^2}$ is the mean value of u^2 then $\overline{u^2}$ can be regarded as being made up of the sum of the squares of the harmonic components. Thus if $\overline{u^2}\,F(n)dn$ is the fraction of $\overline{u^2}$ which is due to frequencies between n and $n+dn$, $\int_0^\infty F(n)dn = 1$

Since each frequency is modified by the filter so that the square of the amplitude is reduced in the ratio $\phi(n) : 1$ it will be seen that Mr Simmons's instrument will record $\overline{u_1^2}$ instead of $\overline{u^2}$, where $\overline{u_1^2} = \overline{u^2}\int_0^\infty F(n)\phi(n)dn$

Mr Simmons has devised a series of filters some of which (high pass) cut out all frequencies below a certain value, and others (low pass) cut out all above this value Considering first the low-pass filters, if $\phi_1(n)$, $\phi_2(n)$ are their characteristics and if $\overline{u_1^2}$, $\overline{u_2^2}$, $\overline{u_3^2}$ are the values of $\overline{u^2}$ measured with the filters in circuit, then

$$\left.\begin{aligned} \frac{\overline{u_1^2}}{\overline{u^2}} &= \int_0^\infty F(n)\phi_1(n)dn, \\ \frac{\overline{u_2^2}}{\overline{u^2}} &= \int_0^\infty F(n)\phi_2(n)dn \text{ etc.} \end{aligned}\right\} \tag{1}$$

The problem is to determine $F(n)$ as a function of n.

Mr Simmons's low-pass filters were so designed that they let nearly the full current through when n was less than some value of N but cut it off entirely when $n > N$, the characteristic function $\phi(n)$ fell rapidly from nearly 1·0 to zero when n passed through a short range in the neighbourhood of N.

As a first approximation therefore we may take N_1 as the value of N at which $\phi = \frac{1}{2}$, and we may assume in the first place

$$\phi_1(n) = 1 \text{ for } n = 0 \text{ to } n = N_1$$

and $$\phi_1(n) = 0 \text{ for } n = N_1 \text{ to } n = \infty.$$

If this were true then the following relationships would be satisfied:

$$\frac{\overline{u_1^2}}{\overline{u^2}} = \int_0^{N_1} F(n)dn, \quad \frac{\overline{u_2^2}}{\overline{u^2}} = \int_0^{N_2} F(n)dn, \tag{2}$$

and evidently $F(n)$ could be determined by graphical differentiation of the curve whose abscissae are N_1, N_2, N_3 and corresponding ordinates

$$\frac{\overline{u_1^2}}{\overline{u^2}}, \quad \frac{\overline{u_2^2}}{\overline{u^2}}, \quad \frac{\overline{u_3^2}}{\overline{u^2}}.$$

Using the measured $\phi(n)$, curves N_1, N_2, N_3 may be determined as described above and the approximate value of $F(n)$ determined If these approximate values are represented by $F_1(n)$ then a further approximation can be obtained as follows. Taking the approximate value $F_1(n)$ determine graphically the value of the integrals

$$\int_0^\infty F_1(n)\,\phi_1(n)\,dn, \quad \int_0^\infty F_1(n)\,\phi_2(n)$$

If it happens that these are all equal to the measured values of $\frac{\overline{u_1^2}}{\overline{u^2}}, \frac{\overline{u_2^2}}{\overline{u^2}}$ then $F_1(n)$ is identical with $F(n)$. In Mr Simmons's case the difference between them was small. Representing this difference by ϵ_1

$$\epsilon_1 = \frac{\overline{u_1^2}}{\overline{u^2}} - \int_0^\infty F_1(n)\,\phi_1(n)\,dn - \int_0^\infty \{F(n) - F_1(n)\}\,\phi_1(n)\,dn \tag{3}$$

It is clear that the same process may be repeated, the equations (3) being treated in the same way as (2), so that

$$\epsilon_1 = \int_0^{N_1} \{F(n) - F_1(n)\}\,dn,$$

$$\epsilon_2 = \int_0^{N_2} \{F(n) - F_1(n)\}\,dn.$$

Thus an approximation to $F(n) - F_1(n)$ may be calculated in the same way as $F(n)$, and since $F_1(n)$ has already been determined, $F(n)$ can be found to a second approximation.

This process can be repeated indefinitely but it is found that one application is sufficient in the case of Mr Simmons's measurements

Comparison of wave-functions for HeH^{++} and HeH^{+}

By C. A. Coulson, Ph.D., *Trinity College, Cambridge*,
and W. E. Duncanson, Ph.D., *Department of Physics, University College, London*

(*Communicated by E. N. da C. Andrade, F.R.S.—Received 14 December 1937*)

1. Introduction

Before trustworthy predictions can be made concerning the reliability of various approximate wave-functions in the case of complicated molecules, it is necessary to study the simplest molecules in as great detail as possible. Such a study enables one to assess the merits and inaccuracies of the different approximations in a way that is impossible with the more complex systems. The simplest of all molecular problems is H_2^+, and this ion has been studied thoroughly by several writers (e.g. Dickinson 1933, Sandemann 1935; Steensholt 1936 *a*, *b*). The simplest two-electron problem is that of H_2, and a very complete knowledge of the wave-functions for this molecule has been obtained (e.g. Weinbaum 1933, Coolidge and James 1933; Coulson 1937*a*). In order of increasing complexity the next molecule is the two-electron ion H_3^+, which has been discussed by Coulson (1935), by Eyring, Rosen and Hirschfelder (1936) and Hirschfelder, Diamond and Eyring (1937). All these molecules are homonuclear, however, so that the binding is predominantly covalent, but the majority of molecules experimentally observed are heteropolar, and then the binding is largely ionic. The present paper, therefore, extends the calculations already made for H_2^+, H_2 and H_3^+, and discusses in detail the two simplest heteronuclear molecules, viz. the ground states of the single-electron ion HeH^{++} and of the double-electron ion HeH^{+}. The object of the paper is primarily to compare the different types of wave-functions, and for this purpose as many diverse methods as possible have been employed, it is not important, from this point of view, that the HeH^{++} ion is unstable and that the HeH^{+} ion is stable. This work may be regarded as the analogue, in the molecular sphere, of a recent paper by Baber and Hassé (1937) on He, in the atomic sphere.

Some of the methods used for the single-electron bond have no immediate counterpart in the discussion of the two-electron bond, and accordingly the two problems are treated independently.

2. THE SINGLE-ELECTRON ION HeH^{++}

In the case of the ion HeH^{++} the problem is that of finding the energy and wave-function of one electron in the presence of two fixed charges $+1$ and $+2$ atomic units (we shall use atomic units a.u. throughout). There are five distinct ways in which this may be done

A. Generalized Morse-Stueckelberg method.

B. Generalized Stark-effect method.

C. LCAO approximation of atomic orbitals

D. Variation expansion in spheroidal co-ordinates.

E. Exact treatment in spheroidal co-ordinates

A. *Generalized Morse-Stueckelberg method*

This method is a development of the work of Morse and Stueckelberg (1929) on the energies of H$_2^+$, and the method applies only when the nuclei are close together, so that we may assume that the ion approximates to the "united-atom" Li^{++}. It should be added that a first approximation to the result of this section has quite recently been obtained by Hassé and Baber (1935) though these authors only used the perturbation method (see below) and did not proceed beyond the first-order correction term.

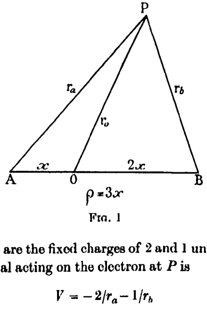


FIG. 1

If A and B (fig. 1) are the fixed charges of 2 and 1 units respectively, then the effective potential acting on the electron at P is

$$V = -2/r_a - 1/r_b$$

If r_0 is the distance OP, where O is the centroid of the positive charges, this may be written

$$V = -3/r_0 + V_{\text{pert}},$$

where

$$V_{\text{pert}} = -2/r_a - 1/r_b + 3/r_0$$

V_{pert} is small if $3x$, the length of AB, is small, and in that case it may be treated as a perturbation. r_0 is measured from the centroid of the positive charges because in this case V_{pert} is of order $1/r_0^2$ at infinity The single-electron orbital is nearly equivalent to an atomic orbital with nuclear charge $+3$ at O, perturbed by V_{pert}. Thus

$$E = E(\text{Li}^{++}) + \int V_{\text{pert}} \,.\, \psi(\text{Li}^{++})^2 \, dv.$$

If we take

$$\psi(\text{Li}^{++}) = (c^3/\pi)^{\frac{1}{2}} e^{-cr_0}, \tag{1}$$

then

$$E = c^2/2 - 3c - cP(t), \tag{2}$$

where $P(t)$ is a known function of the auxiliary variable t, which it is convenient to introduce, and which is defined by

$$t = cx. \tag{3}$$

We may now adopt either the perturbation method or the variation method, as discussed by Coulson (1935, 1937*a*) In the perturbation treatment, we put $c = 3$ for all values of x. This is equivalent to assuming that the fundamental atomic orbital which is perturbed has exactly the wave-function of a Li^{++} ion The energy thus obtained is shown in fig 2 as a function of the nuclear separation $3x$ In the variation treatment, we allow for a stretching of the wave-function due to the perturbation by choosing that value of c in the wave-function (1) which, for a given x, makes the energy a minimum. The energy for such a wave-function, calculated from the usual formula

$$E = \int \psi^* H \psi \, dv \Big/ \int \psi^* \psi \, dv$$

has exactly the same form as that given in equation (2) If we put $\left(\frac{\partial E}{\partial c}\right)_x = 0$, we find that c is given by the equation

$$c = 3 + P(t) + t\frac{dP}{dt}.$$

From this equation a trial value of t is used to calculate the corresponding c, and hence the appropriate nuclear separation ρ which is $3t/c$. The results

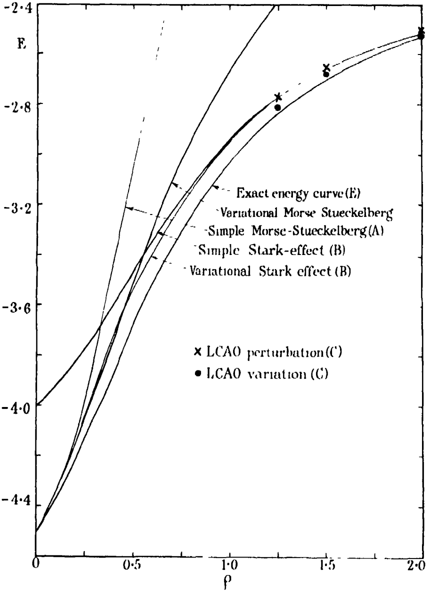


FIG. 2. Electron energy curves of HeH⁺⁺. Various approximations.

of this calculation are shown in fig. 2, and, for comparison with other approximations, in Table I

TABLE I. ELECTRONIC ENERGY AND CONSTANTS OF HeH⁺⁺; VARIOUS APPROXIMATIONS

Section	Method	$\rho = 0·5$	1·0	1·25	1·5	2·0	3·0	4·0
A	Morse-Stueckelberg:							
	(*a*) Perturbation Energy	3·0661	1·2935	0·5851		—	—	—
	(*b*) Variation Energy	3·5080	2·6961	2·4204	2·1967	1·8608	—	—
	c	2·225	1·708	1·543	1·409	1·211	-	—
B	Stark effect							
	(*a*) Perturbation Energy	3·4587	2·9451	2·7811	2·6601	2·4992	2·3333	2·2500
	(*b*) Variation Energy	3·5244	2·9187	2·7819	2·6602	2·4992	2·3333	2·2500
	c	2·326	2·081	2·038	2·017	2·003	2·000	2·000
C	LCAO approximation							
	(*a*) Perturbation Energy	—	-	2·7875	2·6662	2·5023	-	—
	(*b*) Variation Energy			2·8209	2·6817	2·5052	--	—
	c	-		2·135	2·073	2·029		—
	α	—		0·9085	0·8889	0·8157		
D	Variation in spheroidals Energies							
	1	-	2·7769	2·4997	2·2712	1·9355	--	1·2352
	$1, \eta$	–	2·9586	2·7318	2·5404	2·2540	1·8749	1·6206
	$1, \eta, \eta^2$	-	3·0288	2·8296	2·6786	2·4631		1·9627
	$1, \eta, \eta^2, \eta^3$ *		3·0317	2·8371	2·6939	2·5039		2·1321
	$1, \eta, \eta^2, \xi$	-	3·0298	2·8297	2·6786	2·4646		1·9672
	$1, \eta, \eta^2, \xi, \xi^2$	-	3·0300	2·8209	2·6794	2·4648	-	1·9672
	$1, \eta, \eta^2, \xi, \xi^2, \xi\eta$	--	3·0302	2·8209	-	2·4655		1·9731
	$1, \eta, \eta^2, \eta^3, \xi$ *	—	3·0327	—	2·6939	—	—	—
	$1, \eta, \eta^2, \eta^3, \xi, \xi^2$ *	—	3·0330	--	2·6941	-	-	-
	$1, \eta, \eta^2, \eta^3, \xi, \xi^2, \xi\eta$ *	- -	3·0331	-	2·6941	—	--	—
E	Exact energy	3·6655	—		2·6955	—	--	—

B. *Generalized Stark-effect method*

In polar molecules the action of one atom on another to form a molecule is primarily a Stark effect. This is not true with homopolar binding, as the resonance phenomenon shows. At large separations, in the case of HeH⁺⁺, the solitary electron will be centred almost exclusively round the He nucleus, and the energy will be that of a He ion perturbed by an isolated proton. If the wave-function round the He nucleus is taken to be

$$\psi = (c^3/\pi)^{\frac{1}{2}}\, e^{-cr_a}, \tag{4}$$

then the first-order perturbation energy, or, which is the same thing in this case, the energy calculated from the formula

$$E = \int \psi^* H \psi \, dv \Big/ \int \psi^* \psi \, dv,$$

is

$$E = E(\mathrm{He}^+) + \int \frac{1}{r_b} \psi^2 \, dv$$

$$= c^2/2 - 2c - cQ(u), \tag{5}$$

where $Q(u)$ is a known function of the auxiliary variable $u = c\rho$. In the simple Stark effect which corresponds to a perturbation treatment, we assume that ψ is the wave-function of a He ion, so that $c = 2$ (equation (4)) for all ρ, and then

$$E(\rho) = -2 - 2Q(2\rho). \tag{6}$$

The energy obtained with this type of wave-function is shown in fig 2, where it may be compared with the results of other calculations. But there is no reason why the generalization used in the Morse-Stueckelberg method above should not be applied here We use a trial wave-function of type (4), in which c is regarded as a function of ρ, and we minimize the energy, for given ρ, with respect to c This is equivalent to putting $\left(\frac{\partial E}{\partial c}\right)_\rho = 0$ in equation (5), and we find

$$c = 2 + Q(u) + u\frac{dQ}{du} \tag{7}$$

These functions, as in section A, can all be tabulated easily, and the resulting energy curve is shown in fig. 2. Some of the values are also given for comparison in Table 1

C LCAO *approximation of atomic orbitals*

In this approximation, which is applicable at all nuclear distances, and which has been called by Mulliken (1935) the LCAO approximation (linear combination of atomic orbitals) we write for the single-electron molecular orbital:

$$\Psi = \lambda\psi_a + \mu\psi_b, \tag{8}$$

where ψ_a and ψ_b are atomic orbitals round A and B. We shall expect λ to be considerably greater than μ, corresponding to the fact that the electron is mostly round the He nucleus The energy is obtained as a function of λ and μ by the usual formula $E = \int \Psi^* H \Psi dv \Big/ \int \Psi^* \Psi dv$. A general discussion of the energy values with this type of wave-function has been given by Coulson (1937*b*), there are two values of the ratio λ/μ which make the energy a minimum. the upper of these two energies corresponds to an excited state and is to be neglected in favour of the lower one

In the case of HeH^{++} we write

$$\psi_a = (c^3/\pi)^{\frac{1}{2}} e^{-cr_a}, \quad \psi_b = (\alpha^3 c^3/\pi)^{\frac{1}{2}} e^{-\alpha c r_b}. \tag{9}$$

It is convenient to use αc instead of an entirely new parameter in the exponent of ψ_b since now c may be regarded as a scale factor, and the analysis is considerably simplified. The minimum energy is found by variation of both c and α The details of the calculation need not be written down, there is no simple way of minimizing with respect to the parameters, and it is necessary to calculate the energy for values of c and α near to the minimum Whenever ψ_a and ψ_b are used in the rest of this paper, it may be assumed that they have the form given in equation (9). The suffix a will refer to the He nucleus and b to the H nucleus, and the exponents c and α will always be used in the same sense

It is interesting to compare the results of these calculations first when the best possible values of c and α are used, and then when the atomic values $c = 2$ and $\alpha = 1/2$ are used The energies and other constants are shown in Table I where they may be directly compared. It appears, firstly, that the value of α is by no means critical, this is as we should have expected, since α governs the nature of the wave-function near the H nucleus and in this region its amplitude (shown by the ratio μ/λ) is small. Then, secondly, it appears that the value of $c_{\min}$ is slightly greater than the atomic value 2·0, this is an example of the nuclear screening discussed by Coulson (1937*a*) The excess, however, is small, corresponding to the fact that this orbital is only slightly bonding Thirdly, as ρ gets larger, so also the ratio λ/μ increases, corresponding to the fact that as the two positive charges are separated, the solitary electron tends more and more to settle on the He nucleus. The electronic energy curve is shown in fig. 2.

Calculations similar to these have been made by Beach (1936), who, in addition to the above, introduced polarization terms, but his choice of exponents, which restricted α to have the value $\alpha = 1$, is evidently not the best possible, even though it simplifies the calculations. Beach does not give numerical results, and exact comparison therefore is impossible. It will, however, be shown later that the variation of α is much more important in the two-electron problem than in the one-electron problem.

D *Variation expansion in spheroidal co-ordinates*

It is well known that if we use spheroidal co-ordinates ξ, η, ϕ, defined by $\xi = \frac{r_a + r_b}{\rho}$, $\eta = \frac{r_a - r_b}{\rho}$ and ϕ = azimuth, then the wave-equation retains a simple and separable form. Following the method used by Coolidge and James (1933) for H_2, we use a trial wave-function

$$\psi = \Sigma C_{mn} \xi^m \eta^n e^{-\delta\xi}, \tag{10}$$

where C_{mn} and δ are constants chosen so as to minimize the energy. The case of a highly polar orbital such as HeH^{++} is a very unfavourable one for this type of expansion, since it is nearly equivalent to expanding an exponential round one centre in terms of exponentials round another centre, an expansion that converges very slowly.

The first stage is to find the best value for δ. Fig. 3 shows how this depends upon ρ in the case of (*a*) a single-term expansion $e^{-\delta\xi}$, and (*b*) a double-term expansion $e^{-\delta\xi}(1 + a\eta)$. It is unlikely that the addition of more terms would materially alter the best value of δ, and in any case, with more terms, the value of δ is less critical. From these curves, the following table of values was selected for further numerical work.

ρ	1·0	1·25	1·50	2·0	3·0	4·0
δ	1·125	1·375	1·625	2·0	2·75	3·25

That these chosen values of δ were satisfactory is shown by the regular convergence of the variational solutions, according to the rules formulated by Coolidge, James and Present (1936). Table I shows the energy values obtained with particular combinations of terms, thus, for example, the description 1, η, η^2, ξ opposite a wave-function implies that the wave-function was of type

$$\psi = e^{-\delta\xi}\{1 + a_1\eta + a_2\eta^2 + a_3\xi\}$$

The wave-functions marked with an asterisk * all contained a term in η^3, it is evident that without this term only a poor accuracy is attainable, despite the presence of more terms in ξ. An accuracy of 1 : 1000 requires terms up to η^3 in η, but does not even need a term ξ, this is the result of the concentration of charge round one nucleus, a concentration that is governed by the distribution in η. If a further term in η^4 were taken the resulting energy would barely differ from the true one, but, without terms η^3 and η^4, it is not possible to make adequate allowance for the polarity of the orbitals. In the two-electron problem HeH^+, the orbitals are slightly less polar, and we conclude that terms up to and including η^3 are necessary in heteropolar binding. Thus these calculations for heteropolar orbitals are considerably more tedious than for homopolar orbitals, it required, for example, seven terms $(1, \xi, \xi^2, \xi\eta, \eta, \eta^2, \eta^3)$ in the wave-function for HeH^{++} to give as good accuracy as was given by the two-term function $\psi = e^{-\delta\xi}(1+a\eta^2)$ in the case of H_2^+ The situation grows worse as ρ increases, and for values of ρ greater than about 3·0 a.u., we should require a much more complex wave-function, with perhaps as many as twelve terms per electron

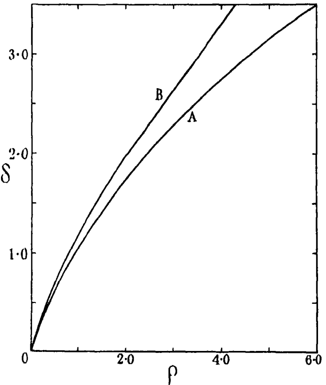


FIG. 3. Best values of δ.

Curve A, $\psi = e^{-\delta\xi}$ Curve B, $\psi = e^{-\delta\xi}(1+a\eta)$

E *Exact treatment in spheroidal co-ordinates*

When there is only one electron present, the wave-equation is separable in ξ and η, the spheroidal co-ordinates of the last section, and it is possible, as a check of the accuracy of the various approximations A–D, to evaluate the energy to any desired degree of approximation, using the method of Hassé and Baber (1935). This separation of co-ordinates is no longer possible when more than one electron is present, but the other methods A–D can still be employed. It is interesting to compare the exact, instead of the relative, accuracies of the various approximations in that one case where it is possible to do so.

If we use equation (18) of the paper by Hassé and Baber, modified by putting $m = 0$, since we are dealing with a σ state, a very rapid approximation is obtained for the ξ equation. This corresponds to the analysis of section D above, where only a few powers of ξ were required in the variational equation. The η equation was more troublesome, and it was found best to rearrange their equation (32) to a simpler form involving only even powers of R_1, and write

$$-\frac{A'}{2} = \frac{u_1}{v_1+}\,\frac{u_2}{v_2+}\,\frac{u_3}{v_3+}\cdots,$$

where

$$u_n = n^2(n^2p^2 - R_1^2),$$

$$v_n = \frac{2n+1}{2}(n^2+n+A').$$

Taking first the case of $\rho = 1{\cdot}5$ a.u., it was found necessary to use eight convergents to get an accuracy of six figures. This indicates that in section D, powers of η up to η^8 are needed to ensure an energy value which is correct to six decimal places. So far as the writers know, this is the only exact discussion of the number of terms needed in a Coolidge-James function for a given accuracy in a particular problem, the usual method is to consider the gain in energy by successive approximations and then estimate (see Coolidge, James and Present 1936) the probable extrapolated limit. It may be mentioned that from Sandemann's discussion of H_2^+ (1935) it is possible to deduce how many terms were needed in that homonuclear case.

The calculation was also made with $\rho = 0{\cdot}5$ a.u., since this provides some exact comparison with the electronic energies calculated by the approximate methods when the nuclei are close together. This time, however, the number of terms required is less; for the wave-function is more nearly spherical and the dependence upon η less important. Instead of eight convergents needed when $\rho = 1{\cdot}5$ a.u. to ensure an energy value correct to six figures, only four are required when $\rho = 0{\cdot}5$ a.u. Again, as when $\rho = 1{\cdot}5$ a.u., only three or four terms of the ξ equation are required.

The results of these calculations are shown in Table I, where they may be compared with the approximate energies of other methods. An exact value for the energy of HeH^{++} is now known for $\rho = 0$, 0·5 and 1·5 a u. By extrapolating from the results of section D, the energy for $\rho = 1{\cdot}0$ and 1·25 can safely be deduced correct at least to 0 002 a u. For ρ greater than 1·5 a u. the energy curve differs insignificantly from $E = -2 - 1/\rho$ It is therefore possible to draw the curve of exact energies, as shown in fig. 2.

3. Discussion of results for HeH^{++}

Several conclusions may be drawn from the numerical results in A–E. Considering first the Morse-Stueckelberg treatment, we notice that in its simplest (perturbation) form, this is only valid over a small range of ρ, even in the more complicated (variational) form, the range of validity, though increased, is still very inadequate, and this method may accordingly be abandoned for more complicated molecules, except in its pictorial and descriptive aspect Thus the general results quoted by Hassé and Baber (1935) or by Bethe (1933) must only be used over a small range of ρ Especially since they use the perturbation rather than the variation method, this range will probably only be from $\rho = 0$ to $\rho = 0{\cdot}25$ a u Nevertheless it is probable that the order of the various levels is correctly given by their formulae, molecular levels such as $2p\sigma$ which have their greatest density around the nuclei are likely to be more correct than levels such as $2p\pi$ where the charge cloud lies away from any nucleus, though this last conclusion might not be true if the variation method were used instead of the perturbation method

Considering next the Stark-effect treatment, this is seen to give a very good energy for values of ρ greater than 2, and this is the case whether or not we use the variational method. At close distances the simple Stark effect is seriously in error, but the variational treatment is remarkably good, considering the crudity of the allowed wave-functions It is probable that this accuracy would be somewhat reduced with less polar orbitals, or where more than one electron participates in the binding, but the inclusion of a term representing the polarization of the larger nucleus would still give a fairly good result. Since methods C and D are much more cumbersome, we conclude that the variational Stark-effect treatment, with allowance for polarization, is the simplest treatment that gives fairly reliable results, though it is least effective in the range $0{\cdot}5 < \rho < 1{\cdot}25$ a.u. Fortunately this range of values is seldom required in practice.

It seems quite possible, from this work, that the Stark-effect treatment might profitably be used to investigate highly polar molecules, such as HCl, for which, at the moment, no suitable type of approximation has been developed. It may be that the best treatment for such problems would be a combination of this method and that used by Buckingham (1937) in calculating Van der Waals' forces between atoms, the method would, of course, be useless in any problems where the binding was nearly, or exactly, homopolar, because it makes no allowance for the resonance effect

The LCAO approximation C is much more cumbersome, due to the presence of quite complicated integrals Variation of the exponent representing the wave-function around the smaller nucleus ψ_b is not very critical, but it is essential to vary the exponent in ψ_a This is unfortunate because it entails considerable labour, especially in the case of more complicated molecules, on the other hand, the result, when the variations have been made, is quite good. The addition of a term representing polarization of the He ion (as e g Dickinson in H_2^+ (1933) or Beach (1936)) would probably give a very accurate energy There is a sense, however, as Dr G. W Wheland pointed out to the authors privately, in which the simple LCAO approximation may be said to allow for the polarization, not just of the molecule as a whole, but also of the larger atom. For if $\Psi = \lambda\psi_a + \mu\psi_b$ is expanded in terms of atomic orbitals round A, the major contribution is $\lambda\psi_a$, but there will be a first-order term $P_1(\cos\theta)$ whose amplitude is proportional to μ, arising from ψ_b, and this term does correspond directly to polarization of the He ion by a uniform field.

Finally, the expansion D in terms of ξ, η, ϕ is able to give any desired accuracy if sufficient terms are taken, but although this is a good approximation for homonuclear problems, it appears to be a bad one for heteronuclear ones, since too many terms in η have to be taken, and the addition of higher powers of η adds much more to the labour, especially when electron-interactions have to be considered. The exact solution E, on the other hand, is impossible except for the one-electron problem, but it does furnish, in that case, a precise standard of accuracy

4 THE TWO-ELECTRON PROBLEM HeH$^+$

The difficulties that hinder an easy solution of the one-electron problem are much enhanced when we proceed to the two-electron problem HeH$^+$, even with the simplest of all molecules H_2 this problem is a serious one. As with the single-electron problem, we shall use as many different methods as possible (F L) and then compare the various results. The methods are:

F. Coolidge-James variation in spheroidals.
G. Ionic wave-function. He atom + proton.
H. LCAO approximation of molecular orbitals.
I. Electron-pair method. Homopolar bond.
J. Electron pair + ionic.
K. Ionic + polar.
L. Ionic + electron pair + polar.

Since we are only considering the ground state, the singlet level $(1s\sigma)^2$, the spin part of the wave-function will separate out and may be omitted, the only restriction upon the space part of the wave-function is that it shall be symmetrical in the co-ordinates of the two electrons.

F. *Coolidge-James variation in spheroidals*

This is a development of the method used for H_2 by Coolidge and James (1933). Unfortunately, owing to the lack of symmetry in HeH^+, terms which do not occur in the wave-function of H_2 do occur here, and thus both the labour and the number of terms are increased. There is a stable minimum of the molecule at about $\rho = 1\cdot5$ a.u., and in view of the labour necessary in this treatment, calculations were only made for $\rho = 1\cdot5$ a.u. The details of these calculations are so similar to those of Coolidge and James that it is not necessary to reproduce them here. A typical term in the wave-function is

$$C_{mnjkp}\, e^{-\delta(\xi_1+\xi_2)}\{\xi_1^m \xi_2^n \eta_1^j \eta_2^k r_{12}^p + \xi_1^n \xi_2^m \eta_1^k \eta_2^j r_{12}^p\},$$

in which r_{12} is the distance between the electrons, and the C_{mnjkp} are constants to be determined by the variation method. The value of δ was taken to be 1·375, rather than 1·625, which was the value used in the one-electron problem, so that the ratio 1·375/1·625 should be nearly the same as the corresponding ratio for H_2 and H_2^+. The results for the various approximations, as the wave-functions grow successively more complex, are shown in Table II. It is clear from this table that higher powers of ξ add very little to the energy, whereas the addition of terms $(\eta_1^3+\eta_2^3)$ and $(\eta_1^4+\eta_2^4)$ would almost certainly yield considerable improvement. The improvement incident upon the inclusion of the r_{12} term, which Coolidge and James found very considerable, is important here, though only about half as important as with H_2. Another four or five terms would be necessary to make the most of this type of wave-function, but in this comparative study the labour for such a calculation did not seem worth while.

If we subtract from the lowest energy value obtained by these calculations the energy of the ground state of the He atom + isolated proton, into which the ion will dissociate at infinite separation, there remains a dissociation energy of 0·012 a.u. = 0·32 volt This figure needs to be corrected for the zero-point energy (about 0·008 a u) but it will still be positive, showing that the ion is stable, and thus confirming the results of Beach (1936). In this approximation there is no difficulty in deciding what is the appropriate energy to subtract when the nuclei are infinitely separated, for when a wave-function of the same type and complexity is used to determine the energy of the ground state of atomic He (Hylleraas 1929) the calculated energy agrees with the observed energy of $-2{\cdot}904$ a u The effect of using more flexible wave-functions for the molecule would be to increase the value 0·32 volt, and the increase would probably be of the order of a volt The normalized wave-function for this approximation, applicable only when $\rho = 1{\cdot}5$ a u., is

$$\Psi = e^{-1\,375(\xi_1+\xi_2)}$$
$$\times \{1{\cdot}744 + 3{\cdot}068(\eta_1+\eta_2) + 2{\cdot}503(\eta_1^2+\eta_2^2) + 1\ 482\,\eta_1\eta_2 + 0{\cdot}7040\,r_{12}\}.$$

G. *Ionic wave-function.* He *atom + proton*

Since the molecule is very ionic in character, the simplest wave-function (one that will be the basis of the remaining wave-functions, being successively modified to allow for polarization and partial formation of a covalent bond, etc) would be

$$\Psi = \psi_a(1)\,\psi_a(2), \tag{11}$$

where, as usual, $\psi_a(1) = (c^3/\pi)^{\frac{1}{2}}\,e^{-cr_{a1}}$.

This wave-function is the molecular analogue of the Stark-effect treatment given in section B for the one-electron problem The energy can be computed from this wave-function by the usual formula and it appears that the molecule, if governed entirely by this wave-function, should be unstable The energy curve, which has no minimum, is shown in fig. 4, and may be compared with the other curves there.

This wave-function is evidently too simple to describe the complicated electron distribution in the molecule, and in the following sections various modifications will be made, which correspond to different chemical assumptions about the nature of the bond.

H. LCAO *approximation of molecular orbitals*

If the molecular orbital for one electron alone (equation (8)) is written $\Phi = \lambda\psi_a + \mu\psi_b$, then the wave-function for the two electrons is

$$\Psi = \Phi(1)\Phi(2). \tag{12}$$

The energy is now obtained by the usual formula as a function of λ/μ, c and α. Some of the integrals that occur in this formula are very troublesome, with one exception, however, they may all be expressed in closed form by the usual methods. The exception is

$$Q = \int \psi_a(1)\,\psi_a(2)\,\psi_b(1)\,\psi_b(2)\,\frac{dv_1\,dv_2}{r_{12}},$$

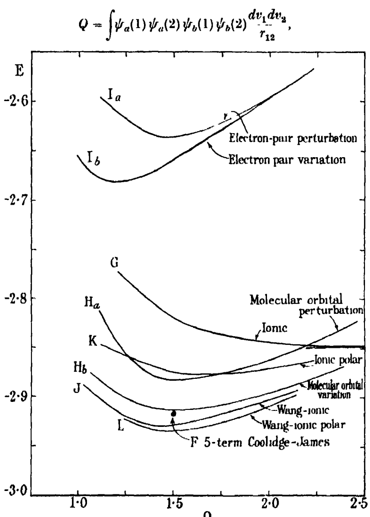


FIG. 4 Energy curves of HeH$^+$. Various approximations

the full calculation of which is given in the appendix It may be worth noting that an attempt to replace this integral by a simpler one that should have approximately the same value was made; by this means an estimate could be obtained of the values of c and α which gave a minimum energy, and it was possible to decide with certainty for which values it was necessary to calculate Q exactly. The integral chosen to replace Q in these preliminary calculations was $\int \psi_a(1)\,\psi_a(1)\,\psi_a(2)\,\psi_b(2)\,\frac{dv_1 dv_2}{r_{12}}$. This latter integral can be calculated in closed form and it appears that, although for large ρ it is considerably too great, yet over a fair range its value lies within 15 % of the true value of Q. In the final variation, however, the values of Q calculated from the formula in the appendix were the only ones used.

The variation with respect to the three parameters λ/μ, c and α, is somewhat laborious and was performed for $\rho = 1{\cdot}25$, $1{\cdot}5$ and $2{\cdot}0$ a.u. The numerical results, after allowing for the Coulomb repulsions of the nuclei, are shown in Table II. Two series of results are shown, first the perturbation method where ψ_a and ψ_b are compelled to have the atomic values ($c = 2, \alpha = 1/2$), and second, the variation method, in which all the parameters were varied.

TABLE II—ENERGIES OF HeH$^+$ CALCULATED BY VARIOUS APPROXIMATIONS

Section	Method	$\rho = 1{\cdot}10$	1·25	1·5	2·0
F	Coolidge-James				
	1			2·3239	—
	$1, (\eta_1+\eta_2)$	—	—	2·7196	—
	$1, (\eta_1+\eta_2), (\eta_1^2+\eta_2^2)$		—	2·9029	—
	$1, (\eta_1+\eta_2), (\eta_1^2+\eta_2^2), \eta_1\eta_2$	—	—	2·9068	—
	$1, (\eta_1+\eta_2), (\eta_1^2+\eta_2^2), \eta_1\eta_2, (\xi_1+\xi_2)$	—	—	2·9072	—
	$1, (\eta_1+\eta_2), (\eta_1^2+\eta_2^2), \eta_1\eta_2, r_{12}$	—	—	2·9161	—
G	He atom + proton		2·7795	2·8190	2·8423
H	LCAO approximation.				
	(*a*) Perturbation		2·8579	2·8834	2·8612
	(*b*) Variation		2·9015	2·9128	2·8938
I	Electron-pair approximation:				
	(*a*) Perturbation		2·6202	2·6358	2·5936
	(*b*) Variation	2·6764	2·6796	2·6576	—
J	Wang-ionic		2·9201	2·9277	2·9011
K	Ionic-polar		2·8582	2·8758	2·8703
L	Ionic-polar-Wang	—	2·9241	2·9327	2·9075

Two Morse curves, shown in fig 4, were put through the calculated points and the constants of the molecule were obtained. These are shown in Table III, together with the constants obtained by other approximations (I–L). The column headed "uncorrected dissociation energy" represents the difference between the calculated energy E of the previous column and the energy at infinite separation of the nuclei, using a wave-function "of the same type" as that used in calculating E. It appears more reasonable to subtract this value of E_{infin} than to subtract the observed energy of $\text{He}+\text{H}^+$; for we desire that the errors in the calculated energy at the equilibrium distance and at infinity shall be as nearly equivalent as possible, a situation to be expected only when wave-functions of the same degree of complexity are used in both calculations (and not always even then!) The final column in Table III shows the dissociation energy corrected for the zero-point vibrational energy.

TABLE III—CALCULATED EQUILIBRIUM CONSTANTS FOR HeH^+

Section	Method	ρ_{equil}	E (a u)	Uncorrected dissociation energy (a u)	Zero-point energy (a u)	Corrected dissociation energy Volts	Corrected dissociation energy a u
H	Molecular-orbital						
	(*a*) Perturbation	1·461	2·8838	0·0362	0·0102	0·70	0·0260
	(*b*) Variation	1·482	2·9128	0·0653	0·0071	1·57	0·0581
I	Electron-pair:						
	(*a*) Perturbation	1·449	2·6365	0·1366	0·0102	3·41	0·1264
	(*b*) Variation	1·197	2·6806	0·1806	0·0114	4·57	0·1692
J	Wang-ionic	1·432	2·9285	0·0808	0·0080	1·97	0·0728
K	Ionic-polar	1·606	2·8769	0·0293	0·0052	0·65	0·0241
L	Ionic-polar-Wang	1·446	2·9332	0·0855	0·0077	2·10	0·0778

One result is nicely verified by these calculations, although it has been known in general terms before λ/μ represents the degree of polarity of the individual orbitals, when there are two electrons instead of one, the repulsion between them will have the effect of throwing more charge on to the H nucleus and will thus make the bond more nearly homopolar, with a value of λ/μ more nearly equal to unity The table below illustrates the change in λ/μ between the molecular orbitals of HeH^{++} and HeH^+.

ρ	λ/μ for HeH^{++}	λ/μ for HeH^+
1·25	5·573	3·099
1·5	7·416	3·501
2·0	10·861	4·613

The best normalized wave-function of this type (12) at the equilibrium distance $\rho = 1\cdot482$ a.u. is

$$\Psi = \Phi(1)\Phi(2),$$

where $$\Phi(1) = 0\cdot85820\psi_a(1)+0\cdot24746\psi_b(1),$$

and $$c = 1\cdot873, \quad \alpha = 0\cdot722.$$

I—*Electron-pair method. Homopolar bond*

The wave-function suitable for use in the electron-pair treatment where we suppose the formation of a covalent bond, is

$$\Psi = \psi_a(1)\,\psi_b(2) + \psi_a(2)\,\psi_b(1). \tag{13}$$

Two cases are considered, first, the perturbation method, which corresponds to the Sugiura treatment of H_2 (1927), and in this the exponents of ψ_a and ψ_b are compelled to take the atomic values $c = 2$ and $\alpha = 1/2$. Second, the variation method, corresponding to the Wang treatment of H_2 (1928), except that two exponents are varied to obtain the minimum energy instead of only one. The calculations in both cases were performed for three values of the internuclear distance and the energies are shown in Table II. With this type of wave-function the energy at infinite separation is that corresponding to $He^+ + H$, since the electrons are restricted to being on different nuclei Morse curves were put through the calculated points and are shown in fig. 4. Table III shows the dissociation energy and equilibrium constants thus deduced

It should be noted that this calculation (with $c = 2$, $\alpha = 1/2$) is the one performed by Glockler and Fuller (1934) These authors, however, did not evaluate the unsymmetrical Sugiura integral, which has been done in the present case The result of including it is to reduce their dissociation energy from 8·1 to about 3·6 volts The perturbation method is clearly inadequate and it is necessary to vary the exponents, in any case, however, it is a poor approximation to consider the bond as being homopolar.

The best normalized wave-function of this type (13), at the equilibrium distance $\rho = 1{\cdot}1970$ a u , is

$$\Psi = 0{\cdot}62308\{\psi_a(1)\,\psi_b(2) + \psi_a(2)\,\psi_b(1)\},$$

where $$c = 2{\cdot}1604, \quad \alpha = 0{\cdot}59214$$

Similar calculations have been made by Beach (1936, p. 355), but here, although c was varied, α was restricted to have the value unity. This lack of flexibility in the wave-function appears to have quite a serious effect upon the energy. Beach's value (estimated from his curve on p 355) at the equilibrium position, is $-2{\cdot}48$ a.u , whereas the present writers' is $-2{\cdot}68$ a.u. The difference between the two values is nearly 6 volts. This is not entirely unexpected, since in the electron-pair treatment one electron may be thought of as centred round the one nucleus and the other round the other nucleus. In this situation their wave-functions should be nearly the appropriate wave-functions for the atoms, and it is unfair to both electrons to make the two exponents the same, with a value 1·72, which is intermediate between the atomic values of 2·0 and 1·0. The error introduced by this approximation would become less if more terms were added to the wave-function (as in Beach 1936) though it would still be appreciable. It is interesting to note that the error is greater in this case than when postulating that the three electrons of a "three-electron bond" have equal exponents in their wave-functions, as in the case of He_2^+ treated by Pauling (1933) and Weinbaum (1935) or in the case of H_2^- treated by Eyring, Hirschfelder and Taylor (1936). A comparison of sections H and I shows that the straightforward molecular-orbital treatment is considerably better than the straightforward electron-pair treatment, a result which has been known in general terms for some time, but not hitherto demonstrated, for a polar molecule.

J *Electron pair + ionic*

The next approximation—one which, as Mulliken (1932) has shown, effectively unites the electron pair and molecular-orbital treatments—is to add ionic terms to the electron pair wave-function. Such a wave-function would be

$$\Psi = \lambda\{\psi_a(1)\,\psi_b(2)+\psi_a(2)\,\psi_b(1)\}+\mu\psi_a(1)\,\psi_a(2)+\nu\psi_b(1)\,\psi_b(2) \qquad (14)$$

The ionic terms $\psi_a(1)\,\psi_a(2)$ correspond to both electrons being on the He nucleus, and $\psi_b(1)\,\psi_b(2)$ to both being on the H nucleus, and the variation is to be performed with respect to c, α and λ μ ν. No appreciable error is introduced if, in the ionic terms, the same exponents are used for ψ_a and ψ_b as in the homopolar terms This approximation corresponds to the treatment of the H_2 molecule by Weinbaum (1933) though, in the H_2 problem, equal weight must be given to both ionic terms. In the case of HeH^+ we may expect the coefficient ν to be very small, and the variation was therefore first performed with $\nu = 0$, so that the problem was that of a normal He atom partly forming a covalent bond with a proton Later, the effect of adding the term $\psi_b(1)\,\psi_b(2)$ was considered.

In the case of $\nu = 0$, we have to minimize with respect to c, α and $\lambda . \mu$; the calculations were made for $\rho = 1\ 25$, 1·5 and 2·0 a.u , and the numerical results, shown in Table II, were used to obtain a Morse curve (fig 4) from which the equilibrium distance and dissociation energy could be computed. These quantities are shown in Table III. It appears, as we should have expected, that the ionic terms are more important than the homopolar terms. The lowering of the energy due to inclusion of the ionic terms is about 0·24 a u. = 6½ volts below the energy of the Wang treatment, and only 0·015 a.u = 0·4 volt below the energy of the molecular-orbital treatment.

This particular approximation has been worked through by Beach with the limiting condition that $\alpha = 1\cdot 0$. This lack of flexibility gives an energy which is 0·023 a.u. = 0·62 volt, too high. In both cases the value of the energy at infinity, obtained by use of an atomic wave-function of the same type, is $-2\cdot 847$ a.u., and this has to be subtracted from the calculated minimum energies to obtain the dissociation energy.

It will be noted that the value of λ/μ becomes less as ρ increases, i.e. the strength of the covalent bond decreases as the nuclei are separated, and both electrons tend to settle on the He nucleus.

The best normalized wave-function of type (14), but with $\nu = 0$, obtained at the equilibrium distance $\rho = 1\cdot 432$ a.u., is

$$\Psi = 0\cdot 65496\psi_a(1)\,\psi_a(2) + 0\cdot 29284\{\psi_a(1)\,\psi_b(2) + \psi_a(2)\,\psi_b(1)\},$$

where
$$c = 1\cdot 9252, \quad \alpha = 0\cdot 7567.$$

The wave-function was next considered, in which ν is allowed to vary to give the lowest energy. Since the energy corresponding to $\psi_b(1)\,\psi_b(2)$ is very high relative to the other components of the wave-function (14), we shall not expect a large contribution from this function. Two particular cases were worked out in full, minimizing with respect to λ, μ, ν for fixed values of c and α. The results were

ρ	α	c	$E_{\nu=0}$	$E_{\nu\neq 0}$
1·25	0·75	1·92	$-2\cdot 9158$	$-2\cdot 9163$
2·00	0·75	1·75	$-2\cdot 8989$	$-2\cdot 8997$

Thus the gain in energy is only 0·0005 a.u. or 0·013 volt in the one case, and 0·0008 a.u., or 0·022 volt, in the other. This additional stability is very small and, in view of its smallness, it was not considered worth while to carry the calculations further by varying the exponents. It appears that, at the equilibrium distance, the lowering of the energy value due to the inclusion of this term may be taken to be about 0·018 volt = 0·0007 a.u. Even if we omit the term completely, the error will be less than that which is inherent in any wave-function that lacks an r_{12} term in its expansion.

K. *Ionic + polar*

In this treatment we suppose that the effect of the H nucleus on the He atom is to polarize it; this polarization may be described by saying that there is a finite probability that one of the electrons shall be in a polar orbital, defined by

$$\phi_a = (d^5/\pi)^{\frac{1}{2}}\, r_a \cos\theta_a\, e^{-dr_a}. \tag{15}$$

The appropriate wave-function to take for the molecule is, therefore,

$$\Psi = \mu\psi_a(1)\,\psi_a(2) + \nu\{\phi_a(1)\,\psi_a(2) + \phi_a(2)\,\psi_a(1)\}. \tag{16}$$

There is no need to consider the possibility that both electrons are in polar orbitals ϕ, since the energy corresponding to such a wave-function is too great, and the matrix components with the other component wave-functions would be negligibly small. The inclusion of this polar term in molecular problems is due to Rosen (1931 *a*) who used wave-functions of type (15) in his study of H_2. The energy resulting from wave-function (16) was minimized with respect to c, d and ν/μ and the results are shown in Table II. A Morse curve (fig 4) was put through the calculated points, and the equilibrium constants determined by this means are given in Table III. It is seen that stability for the molecule is achieved, and here there is no reference to a homopolar bond At the equilibrium distance $\rho = 1{\cdot}6055$ a u., and the best normalized wave-function of this type is

$$\Psi = 0{\cdot}98267\,\psi_a(1)\,\psi_a(2) + 0{\cdot}13107\{\psi_a(1)\,\phi_a(2) + \psi_a(2)\,\phi_a(1)\},$$

with $$c = 1{\cdot}7126 \quad \text{and} \quad d = 1{\cdot}5586$$

L *Ionic + electron pair + polar*

The final approximation that we discuss is one in which allowance is made both for covalency and for polarization, this corresponds to the treatment of H_2 given by Weinbaum (1933) Using the polar orbital (15), the wave-function is

$$\Psi = \lambda\{\psi_a(1)\,\psi_b(2) + \psi_a(2)\,\psi_b(1)\} + \mu\psi_a(1)\,\psi_a(2) + \nu\{\psi_a(1)\,\phi_a(2) + \psi_a(2)\,\phi_a(1)\}. \tag{17}$$

There was some difficulty in deciding whether it would be easier to give the exponent d in ϕ the value c or $c\alpha$ The value $c\alpha$ was chosen because that simplified the calculations of some of the exchange integrals. These calculations were made for $\rho = 1{\cdot}25$, $1{\cdot}5$ and $2{\cdot}0$ a.u , and the energy was minimized with respect to c, α and $\lambda : \mu : \nu$ Table II shows the results, and in fig. 4 the Morse curve is drawn through the calculated points The resulting equilibrium constants are given in Table III. It is interesting to note that ν/μ, which shows the importance of the polarization terms, has a maximum value when the molecule is in the equilibrium configuration, showing that the polarization forces are, in proportion, most important just when they are most useful in stabilizing the molecule.

This wave-function has been used by Beach with the usual limitation $\alpha = 1$, the energy thus obtained lies $0{\cdot}011$ a.u $= 0{\cdot}30$ volt above the energy without this limitation. The gain of energy due to inclusion of the polar terms is only 0·127 volt in the writers' wave-function, but it is 1·05 volts in Beach's case. The large gain obtained by Beach is due to a bad choice of exponents in the homopolar wave-function, and the fact that inclusion of polar terms gives greater flexibility and allows the wave-function to adapt itself better to the limitation of bad exponents.

The energy of dissociation given by the wave-function of this section is 2·10 volts, if the energy at infinite separation is taken to be that of a He atom ($-2{\cdot}847$) calculated from a similar type of wave-function. Unfortunately this atomic wave-function has not so much flexibility as the molecular wave-function, and its energy will probably be more in error than the molecular energy, so that it is probable that the true dissociation energy is rather less than 2·10 volts. A lower limit can be obtained if we subtract from the calculated value ($-2{\cdot}9332$ a.u.) the observed energy of a normal He atom ($-2{\cdot}9035$ a.u.). The dissociation energy then becomes 0·0297 a.u., or with allowance for zero-point energy, 0·0220 a.u., or 0·594 volt. To this we should add about 0·018 volt, representing the contribution of the ionic term with both electrons on the H nucleus (section J) and this gives a lower limit for the dissociation energy of 0·612 volt. The true value is probably between 1 and 1·5 volts.

The value of the fundamental vibration frequency obtained from the Morse curve, is $\tilde{\nu} = 3380\,\mathrm{cm}^{-1}$. It may be compared with the value given by Beach, which is $2800\,\mathrm{cm}^{-1}$. Beach's value was obtained by putting a parabola through three points near the minimum, whereas the present writers used a Morse curve. If the three points were not very close together and near the minimum, the difference in the two estimates would easily be attributable to this difference in technique and, in fact, the direction of the difference supports this conclusion.

The best normalized wave-function of this type, at the equilibrium distance $\rho = 1{\cdot}446$ a.u., is

$$\Psi = 0{\cdot}71730\,\psi_a(1)\,\psi_a(2) + 0{\cdot}24843\{\psi_a(1)\,\psi_b(2) + \psi_a(2)\,\psi_b(1)\} + 0{\cdot}02979\{\psi_a(1)\,\phi_a(2) + \psi_a(2)\,\phi_a(1)\},$$

where $c = 1{\cdot}8900$ and $\alpha = 0{\cdot}80560$

5 DISCUSSION

The work recorded in sections F–L provides an interesting comparison of the various methods. Taking first the simple molecular-orbital and electron-pair treatments, it is obvious from Table III that from every point of view the molecular-orbital treatment is better than the electron-pair treatment. The internuclear distance and zero-point energy are substantially correct and the total energy appears to be only about 0·5–1·0 volt too high Taken in conjunction with the work of Coulson (1937*a*) on the homopolar molecule H_2, where it was found that the orbital treatment was only slightly inferior to the pair treatment, this means that the molecular-orbital approximation is established as, on the whole, a more accurate quantitative approximation than the other The electron-pair wave-function relies for its binding mainly upon the exchange term, and this, even in homopolar bonds, has a rapid decrease with increase of internuclear distance, in heteropolar bonds the decrease is still more rapid. The molecular-orbital wave-function, on the other hand, takes polarization and ionicity into account (generally allowing them too large an emphasis) and these forces are operative over greater distances Thus it comes about that in the case of HeH^+ where the bond is polar, the pure electron-pair wave-function yields far too low an internuclear distance, and the pure molecular orbital-wave-function a value slightly too high. It is interesting to note that precisely the opposite effect is noticed (see Coulson 1937*a*) in the homopolar bond H_2

Another conclusion from this work is that there is no hope of obtaining a good value for the energy unless a great deal of flexibility is allowed in the wave-function, and this is as important in the case of polar binding as in the case of a covalent link In homopolar molecules there are considerable computational difficulties, in heteropolar molecules the position is worse. It is, of course, essential to use a variational method, and if atomic orbitals are used as components in the final wave-function (as in G–L), then all the exponents should be varied The effect of not varying them all, as comparison with the results of Beach shows, is to introduce errors that may vary from ½ to 6 volts It is interesting to exhibit the errors that accrue, in this case, through making the limitation that $\alpha = 1$, and Table IV reveals the fact that as greater flexibility is allowed in the wave-function, the particular values that are chosen for the various parameters become less and less critical. This result had already been noticed by Coolidge and James (1933) in their discussion of H_2.

One other interesting deduction can be made from this work. It has become customary to refer to the strength of a particular bond as being partly homopolar, and partly polarization Thus Beach (1936, p. 357) refers to the HeH^+ bond "It is qualitatively correct to say that two-thirds of the stability is due to the formation of a covalent bond, and one-third is due to polarization of the He atom." It is possible to make this a little clearer from our results. Thus, measuring all energies at the internuclear distance

1·446 obtained in the last section L, we find that the energy of the purely ionic wave-function G is $-2·816$ a.u. The energy of a wave-function J which is ionic + homopolar bond, is $-2·927$ a.u., and that of a wave-function K which is ionic + polarization, is $-2·872$ a.u. It follows that the gain in energy beyond that of a normal He atom, due to inclusion of homopolar binding, is 0·111 a.u., and due to polarization of the He atom, is 0·056 a.u. We may therefore say that the bond, at the equilibrium distance, is almost exactly *one-third* polarization and two-thirds homopolar. It is interesting that our results agree so closely with Beach's, in view of the large difference in the energy of the homopolar part of the wave-function.

TABLE IV—ERROR IN ENERGY DUE TO LACK OF VARIATION OF EXPONENT α

Section	Method	Min. energy with limitation $\alpha = 1$	Min. energy without limitation $\alpha = 1$	Difference a.u.	Difference Volts
I.	Wang	$-2·486$*	$-2·6806$	0·20	5·4
J.	Wang-ionic	$-2·9025$	$-2·9285$	0·0260	0·70
K.	Ionic-polar	$-2·8750$	$-2·8769$	0·0019	0·05
L.	Ionic-polar-Wang	$-2·9220$	$-2·9332$	0·0112	0·30

* Estimated from Beach's curve.

In conclusion, the writers would like to thank Professor Andrade for putting at their disposal, from time to time, a calculating machine which was originally purchased by a Government grant from the Royal Society.

SUMMARY

As many different approximations as possible have been used in a comparative study of the wave-functions of HeH^{++} and HeH^{+}. In the latter ion, the molecular-orbital approximation is found to be much better than the electron-pair approximation. Ionic terms and polar terms are included in the final wave-function, and an expansion is given in terms of spheroidal co-ordinates The lowest value obtained for the energy of HeH^{+} is $-2·935$ a.u., with an internuclear distance of 1·446 a.u. = 0·764 A°. The dissociation energy certainly lies within 0·61 and 2·10 volts, probably about 1·5 volts. The fundamental vibration frequency is 3380 cm.$^{-1}$.

APPENDIX

The only integral used in this work that cannot be evaluated in closed form by conversion to spheroidal co-ordinates ξ, η, ϕ or whose value has not been given in the literature, is the integral

$$Q = \int \frac{1}{r_{12}} e^{-cr_{a1}-cr_{a2}-c\alpha r_{b1}-c\alpha r_{b2}} \, dv_1 \, dv_2.$$

This integral must be evaluated in the form of an infinite series. Using the well-known Neumann expansion of $1/r_{12}$ in terms of ξ, η, ϕ (see, e.g., Rosen 1931 *b*), and writing

$$\gamma_1 = c\rho(1+\alpha)/2, \quad \gamma_2 = c\rho(1-\alpha)/2,$$

the integral becomes

$$Q = 4\pi^2(\rho/2)^5 \sum_{\tau=0}^{\infty} (2\tau+1)$$
$$\times \{[G^0_\tau(\gamma_2)]^2 H_\tau(2.2.\gamma_1) - 2G^0_\tau(\gamma_2)\, G^2_\tau(\gamma_2) H_\tau(2.0.\gamma_1) + [G^2_\tau(\gamma_2)]^2 H_\tau(0\ \ 0.\gamma_1)\},$$

H and *G* are the functions defined by Rosen (1931 *b*) and Coolidge and James (1933), and they may be written:

$$H_\tau(m.n.\gamma) = \int_1^\infty \int_1^{\xi_1} Q_\tau(\xi_1)\, P_\tau(\xi_2)\, \xi_1^m \xi_2^n e^{-\gamma(\xi_1+\xi_2)} \, d\xi_2 \, d\xi_1$$
$$+ \int_1^\infty \int_1^{\xi_2} Q_\tau(\xi_2)\, P_\tau(\xi_1)\, \xi_1^m \xi_2^n e^{-\gamma(\xi_1+\xi_2)} \, d\xi_1 \, d\xi_2.$$

$$G^p_\tau(\gamma) = \int_{-1}^{+1} \eta^p P_\tau(\eta)\, e^{-\gamma\eta} \, d\eta.$$

Recurrence relations are available (Coolidge and James 1933) for the *H* functions; the *G* functions were evaluated by means of the two equations:

$$(2\tau+1)\, G^0_\tau(\gamma) = \gamma\{G^0_{\tau+1} - G^0_{\tau-1}\},$$

$$(2\tau-1)(2\tau+1)(2\tau+3)\, G^2_\tau = \tau(\tau-1)\, G^0_{\tau-2}$$
$$+ (4\tau^3+6\tau^2-1)\, G^0_\tau + (\tau+1)(\tau+2)\, G^0_{\tau+2}.$$

It should be noted that although Q is given as an infinite series, the successive terms rapidly become small. Thus, for the values used in the calculations of this paper, it is sufficient to take terms up to $\tau = 3$ to get an accuracy of at least 1 in 10^5, which is equivalent to obtaining the energy represented by this integral with an error not greater than 0·0003 e-volt. This is quite sufficient for our purposes, though the addition of one more term would increase the accuracy about twenty times.

REFERENCES

Baber and Hassé 1937 *Proc. Camb Phil. Soc.* **33**, 253.
Beach 1936 *J Chem Phys.* **4**, 353.
Bethe 1933 "Handbuch der Physik", **24**, 527.
Buckingham 1937 *Proc. Roy. Soc.* A, **160**, 94.
Coolidge and James 1933 *J Chem. Phys* **1**, 825.
Coolidge, James and Present 1936 *J. Chem. Phys.* **4**, 187.
Coulson 1935 *Proc. Camb. Phil Soc* **31**, 244.
— 1937*a* *Trans. Faraday Soc* **33**, 1478.
— 1937*b* *Proc. Camb. Phil Soc.* **33**, 111.
Dickinson 1933 *J. Chem. Phys* **1**, 317
Eyring, Hirschfelder and Taylor 1936 *J. Chem. Phys.* **4**, 479
Eyring, Rosen and Hirschfelder 1936 *J. Chem. Phys* **4**, 121, 134.
Glockler and Fuller 1934 *J. Chem Phys.* **1**, 886
Hassé and Baber 1935 *Proc Camb. Phil Soc.* **31**, 564.
Hirschfelder, Diamond and Eyring 1937 *J. Chem. Phys.* **5**, 695.
Hylleraas 1929 *Z. Phys.* **54**, 347
Morse and Stueckelberg 1929 *Phys Rev.* **33**, 932.
Mulliken 1932 *Phys. Rev.* **41**, 49.
— 1935 *J. Chem Phys.* **3**, 375.
Pauling 1933 *J Chem. Phys* **1**, 54
Rosen 1931*a* *Phys Rev.* **38**, 2009.
— 1931*b* *Phys Rev.* **38**, 255
Sandemann 1935 *Proc Roy Soc. Edin.* **55**, 72.
Steensholt 1936*a* *Z. Phys.* **100**, 547.
— 1936*b* *Avh. Norske VidenskAkad.* no. 1.
Sugiura 1927 *Z. Phys* **45**, 484.
Wang 1928 *Phys Rev.* **31**, 579.
Weinbaum 1933 *J Chem Phys* **1**, 593
— 1935 *J. Chem Phys.* **3**, 547.

The production of gamma-rays by neutrons

By E. H. S. Burhop, B.A., M.Sc., R. D. Hill, M.Sc. and
A. A. Townsend, M.Sc.

Natural Philosophy Laboratory, University of Melbourne

(*Communicated by Professor T. H. Laby, F.R.S.*)

Introduction

The production of γ-radiation due to the passage of neutrons through matter was discovered by Lea (1935) and has been subsequently investigated by a number of workers. The experiments of Lea with fast neutrons appeared to indicate that the γ-radiation arises from a process of inelastic scattering of neutrons. Evidence for the occurrence of inelastic scattering of fast neutrons has been obtained by Danysz and others (1934) and Ehrenberg (1935) and it has been pointed out by Bohr that on current theories of the interaction between fast neutrons and atomic nuclei most of the scattering would be expected to be inelastic. On the other hand, measurements of the absorption of slow neutrons in elements which emit γ-rays (e.g. Cd, Hg) show that the absorption is a sharp resonance process, and it is assumed that the γ-radiation in these cases arises from neutron capture. The present work was undertaken with a view to determining the energies of the neutrons which give rise to the γ-radiation, in an attempt to throw some light on the mechanism of its production.

Method of estimating the energy of the neutrons

The energy of homogeneous neutron groups can be estimated by measuring their absorption coefficient in boron. This method has been applied to estimate the energy of the neutrons which give rise to artificial β-radioactivity in the case of a number of nuclei, but has not been applied previously to the estimation of the energy of neutron groups responsible for the production of γ-radiation.

Validity of the $1/v$ law for boron absorption

The boron absorption method of determining neutron energies rests on the assumption that the absorption coefficient of slow neutrons in boron is inversely proportional to the neutron velocity. This law, which follows from very general theoretical considerations, has been experimentally verified in the region of thermal energies by Rasetti and co-workers (1936), using a mechanical method to vary the relative velocity of the neutrons and the absorbing material.

In energy regions higher than the thermal the evidence for the law is mainly indirect. Thus Goldsmith and Rasetti (1936) have measured the energies of those neutrons which excite β-ray activities in various elements by carrying out absorption measurements in both lithium and boron. Assuming the $1/v$ law to hold for both lithium and boron absorption, the values of these energies come out to be the same in either case. It follows from this that either the $1/v$ law is valid for both boron and lithium absorption or else both of these absorbers depart from the law in the same manner.

Further, Preiswerk and von Halban (1936) (and independently Amaldi and Fermi (1936)) have described another method for the determination of the energies of neutron resonance absorption groups quite independent of the $1/v$ absorption law. The measurements of the energies of the groups obtained by this method show approximate agreement with those made assuming the validity of the $1/v$ law

The $1/v$ method of investigating the energies of neutrons responsible for the production of γ-radiation involves

(1) The determination of the absorption coefficient in boron for thermal neutrons. The mean energy of these neutrons can be calculated on the assumption of thermal equilibrium between the neutrons of this group and the paraffin molecules used to reduce their velocity.

(2) The determination of boron absorption curves for the neutrons which excite γ-radiation in the different materials being investigated.

From the former determination the boron absorption coefficient for neutrons of a certain energy can be obtained; and assuming the $1/v$ law the energy of the neutron groups responsible for the different activities can be obtained from the latter.

EXPERIMENTAL ARRANGEMENT

Fig. 1 shows the arrangement of apparatus used. The γ-radiation was detected by means of a Geiger-Müller counter 6 cm. long, 3 cm. diameter, enclosed in a cylinder of lead with a wall sufficiently thick (2 mm.) to prevent any β-particles arising in the specimen from reaching the counter. The specimens in which the γ-radiation was induced were in two forms, viz.

(1) In the case of Ag, Cd, I, Hg hollow cylinders surrounding the counter.

(2) In the case of Sb and As rectangular boxes 9 × 7 cm., of appropriate thickness to produce a measurable effect and placed directly beneath the counter.

In either case screens of B_4C were placed above and around the sides of the specimen to absorb slow neutrons scattered from neighbouring objects. This precaution was found to be necessary since the effect with silver was reduced by about 20 % when the screens were employed.

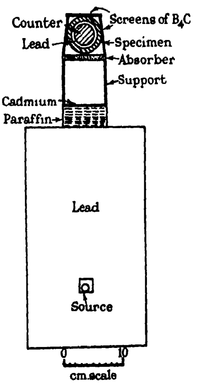


FIG. 1 Experimental arrangement.

MAGNITUDE OF OBSERVED EFFECTS

With the counter used the normal cosmic ray count was approximately 60 per min. The introduction of a 350 millicurie Rn + Be neutron source increased the count to 220 per min. When a cylinder 7 cm. long, 4 mm. thick of silver was placed round the lead cylinder containing the counter the counting rate was increased further to 300 per min.

DETERMINATION OF THE ABSORPTION COEFFICIENT OF THERMAL NEUTRONS IN BORON

In general the measured value of the absorption coefficient of thermal neutrons will depend upon the particular type of neutron detector used. Previous workers have used a detector for which the sensitivity is inversely proportional to the neutron velocity. Under these conditions Bethe and Placzek (1937) have shown that the measured absorption coefficient in boron corresponds to that for neutrons having a mean energy of $\pi kT/4$, T being the temperature of the paraffin in which the neutrons have been reduced to thermal energies.

In our experiments the arrangement of fig. 1 was used. A cylinder of cadmium 0·5 mm. thick and 7 cm. long fitting tightly round the lead cylinder in which the counter was placed was used as a neutron detector. The boron in the form of boron carbide was contained in thin soda-glass containers 9×7 cm. placed directly beneath the specimen in which the γ-radiation was being produced. Each container had the same thickness, the boron content being varied by mixing the B_4C with powdered carbon in different proportions in the various containers Appropriate corrections were made for the absorption of γ-radiation arising from the radon source in the specimen and absorbers.

In this method of experiment the neutron detector registers all thermal neutrons independently of their velocity. In these circumstances it can be shown that the measured boron absorption coefficient corresponds to that for neutrons of energy $4kT/\pi$. On account of this it is to be expected that the absorption coefficient obtained from this measurement should be lower than that obtained by other workers in the ratio $\pi/4$ (assuming the $1/v$ law to be valid).

Fig. 2 shows the boron absorption curve obtained by this method. The value of the mass absorption coefficient deduced from the early part of this curve is $\mu/\rho = 30 \pm 2$ cm.^{2} gm.$^{-1}$. The curve is not a true exponential owing to the gradual hardening of the neutron beam in passing through the boron. The form of the curve agrees well with that to be expected assuming a $1/v$ law for neutron absorption in boron

The absorption coefficient of thermal neutrons in boron has also been determined using as detector the γ-rays induced in a silver cylinder of the same dimensions as the cadmium and correcting for the production of γ-radiation in the specimen by neutrons of non-thermal energies. In the case of silver the absorption coefficient for thermal neutrons is much smaller than in the case of cadmium and only a small fraction of incident thermal neutrons will be absorbed in a silver specimen of this thickness, so that the sensitivity of detection is approximately proportional to $1/v$. The mean energy of thermal neutrons absorbed by a thin layer of boron is then $\pi kT/4$. Assuming the $1/v$ law for boron absorption one would then expect a value of the mass absorption coefficient in this case of $\mu/\rho = 4/\pi \times 30$ cm.^{2} gm.$^{-1}$ = 38 cm.^{2} gm.$^{-1}$. The value actually obtained was 35 ± 5 cm^2 gm.$^{-1}$

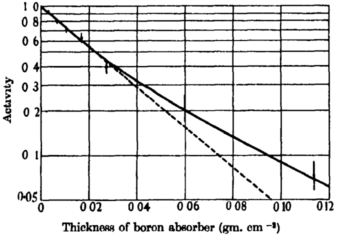


FIG. 2 Absorption in boron of neutrons which excite γ-rays in cadmium.

Table I shows the values of μ/ρ so far obtained by different methods of detection.

TABLE I. BORON ABSORPTION COEFFICIENTS FOR THERMAL NEUTRONS

Observer	Method of detection used	μ/ρ for neutrons of energy $\pi kT/4$ ($T = 290°$ K.)
Mitchell (1936)	Boron ionization chamber	27·7 gm.$^{-1}$ cm.^{2}
Goldsmith and Rasetti (1936)	Rhodium β-radioactivity	28·0 ,,
Amaldi and Fermi (1936)	Rhodium β-radioactivity	38·0 ,,
Livingston and Hoffman (1937)	Boron ionization chamber	36·0 ,,
Authors	Cadmium γ-radiation	*38·0 ± 2·0 ,,
Authors	Silver γ-radiation	*35·0 ± 5·0 ,,

* Calculated from measured absorption coefficient for neutrons of energy $4kT/\pi$.

It is seen that there is considerable divergence between the values obtained by different observers. Our measurements are in good agreement with those of Amaldi and Fermi and of Livingston and Hoffman.

In this paper the value $\mu/\rho = 30$ cm.^{2} gm.$^{-1}$ has been taken for the boron absorption coefficient for neutrons of energy 0·033 V ($= 4kT/\pi$ if $T = 290°$ K.),

DETERMINATION OF THE BORON ABSORPTION CURVE FOR NON-THERMAL NEUTRONS WHICH EXCITE γ-RADIATION

In a manner exactly similar to that described in the last section boron absorption curves have been obtained for the non-thermal neutrons which excite γ-radiation in the specimens tabulated below.

TABLE II. THICKNESS OF SPECIMENS USED

Specimen	Mass/unit area
Silver	4 2 gm cm $^{-2}$
Silver	0 525 ,,
Arsenic	3 61 ,,
Antimony	5 88 ,,
Iodine	5·5 ,,
Mercury	11·0 ,,

With the exception of one of the silver specimens the above are considerably thicker than detectors used in the case of β-ray measurements. Specimens of this thickness were necessary in order to get effects of a convenient magnitude with the sources available to us. Further, they yield information about neutrons which are captured into a considerable number of resonance levels in contrast with curves obtained with thin indicators which give information about the resonance level of lowest energy only.

Figs. 3 and 4 show the curves obtained for the γ-ray activity of the specimens as a function of the thickness of the boron absorber. All the curves exhibit the same general features, viz a rapid initial decrease of intensity with boron thickness followed by a region in which the intensity decreases very slowly.*

THEORETICAL FORM OF THE ABSORPTION CURVES

Bethe and Placzek (1937) have shown how the boron absorption curves can be analysed along the lines of the theory of neutron capture of Bohr (1936) and of Breit and Wigner (1936) when thin specimens are used as neutron detectors. The theory is here extended to the case of detectors of finite thickness. We calculate first the activity produced in a specimen of thickness l due to the absorption of neutrons into a single resonance level.

Bethe and Placzek deduce for the activity of a thin specimen due to

---

* The thick-silver curve of fig 3 is slightly different from that previously reported by the authors (1936) Previously, insufficient correction was made for the absorption of γ-rays from the radon source in the boron containers.

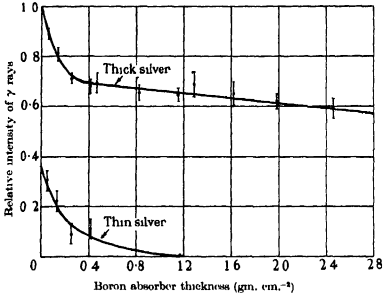


FIG. 3. Absorption in boron of neutrons which excite γ-rays in two specimens of silver. The thicknesses are given in the text The ordinates for zero boron absorber thickness represent approximately the relative activity.

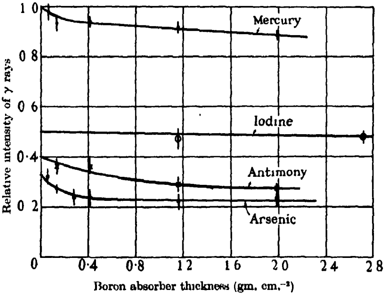


FIG. 4. Absorption in boron of neutrons which excite γ-rays in mercury, iodine, antimony and arsenic. The ordinates for zero boron absorber thickness represent approximately the relative intensity of the activities for detectors of thickness given in the text.

capture into a level g of the intermediate state (produced by neutrons which have reached thermal equilibrium with the paraffin used to slow them down)

$$A_g = C\delta\Gamma'_n E_g^{-\frac{1}{2}}, \tag{1}$$

where E_g is the energy of the level g, δ the thickness of the specimen, C a constant depending on the rate of emission of neutrons from the source and their mean free path in the paraffin used to slow them down.

Γ'_n, the so-called "reduced neutron width", is given by $\Gamma'_n = \Gamma_n E_g^{-\frac{1}{2}}$, Γ_n being the width of the resonance level into which the capture occurs arising from the possibility of the emission of a neutron from that level We shall assume, with Bethe and Placzek, that there is little variation in Γ'_n between the different levels for the same nucleus. We assume further that the neutrons are absorbed exponentially with absorption coefficient $\alpha E_g^{-\frac{1}{2}}$ in the material of the detector The assumption of exponential absorption is not strictly correct on account of the finite width of the level. From equation (60) of Bethe and Placzek's paper it is seen that the assumption of an absorption coefficient proportional to $E_g^{-\frac{1}{2}}$ is equivalent to assuming Γ the total width of the levels of the same nucleus to be almost constant. Fortunately it is found that the particular value chosen for the neutron absorption coefficient in the specimen does not greatly affect the conclusions to be drawn, so that these assumptions are not critical.

The activity produced in a detector of thickness l is

$$A_g = C\Gamma'_n E_g^{-\frac{1}{2}}\int_0^l e^{-\alpha x E_g^{-\frac{1}{2}}} e^{-\kappa(l-x)}\,dx$$

$$= \frac{C\Gamma'_n(e^{-\kappa l} - e^{-\alpha l E_g^{-\frac{1}{2}}})}{E_g(\alpha - \kappa E_g^{\frac{1}{2}})}, \tag{2}$$

where κ is the absorption coefficient of the γ-radiation in the specimen itself.

If before reaching the detector the neutrons are filtered through a layer of boron of thickness t the activity produced becomes approximately

$$A_g(t) = \frac{C\Gamma'_n e^{-\beta t E_g^{-\frac{1}{2}}}(e^{-\kappa l} - e^{-\alpha l E_g^{-\frac{1}{2}}})}{E_g(\alpha - \kappa E_g^{\frac{1}{2}})}, \tag{3}$$

where $\beta E_g^{-\frac{1}{2}}$ is the absorption coefficient of neutrons of kinetic energy E_g in boron

We suppose now that the levels are uniformly spaced, D volts being the spacing between consecutive energy levels and E_1 the energy of the lowest level.

Then $$E_g = E_1 + (g-1)D. \tag{4}$$

The total activity $A(t)$ is then given by the summation $\sum_{g=1}^{\infty} A_g(t)$. Transforming the sum into an integral,

$$A(t) \simeq \frac{A_1(t)}{2} + C\Gamma'_n \int_1^{\infty} \frac{e^{-\beta t E_g^{-\frac{1}{2}}}(e^{-\kappa l} - e^{-\alpha l E_g^{-\frac{1}{2}}})}{E_g(\alpha - \kappa E_g^{\frac{1}{2}})} dg$$

$$= \frac{C\Gamma'_n}{E_1} \left\{ \frac{e^{-\beta t E_1^{-\frac{1}{2}}}(e^{-\kappa l} - e^{-\alpha l E_1^{-\frac{1}{2}}})}{2(\alpha - \kappa E_1^{\frac{1}{2}})} \right\} + \frac{2E_1}{D} e^{-\frac{(\alpha l + \beta t)\kappa}{\alpha}} \left[Ei\left\{ -\beta t \left(E_1^{\frac{1}{2}} - \frac{\kappa}{\alpha} \right) \right\} \right.$$

$$\left. - Ei\left(\frac{\beta t \kappa}{\alpha} \right) - Ei\left\{ -(\alpha l + \beta t)\left(E_1^{\frac{1}{2}} - \frac{\kappa}{\alpha} \right) \right\} + Ei\left\{ (\alpha l + \beta t)\frac{\kappa}{\alpha} \right\} \right],$$

where $$Ei(x) = \int_{-\infty}^{x} \frac{e^u du}{u}. \tag{5}$$

Expression (5) gives the activity in a convenient form for comparison with the experimental results, and shows particularly well the variation of the form of $A(t)$ with the ratio D/E_1. For $D/E_1 = \infty$, the first term alone gives the shape of the boron absorption curve, whilst for $D/E_1 = 0$ the second term gives the shape of the curve For finite values of D/E_1 the appropriate curve lies between these two extremes.

APPLICATION OF THE THEORY TO THE EXPERIMENTAL CURVES

The above theory has been applied to the analysis of the experimental boron absorption curves for neutrons which excite γ-rays in silver. In order to apply equation (5) it is necessary to know the values of certain constants concerning the absorption of neutrons in boron and silver and of γ-rays in silver.

The absorption coefficient κ of silver γ-rays in silver is calculated as 0·035 cm.^{2} gm. ^{1}, assuming the quantum energy of these radiations to be 4 million electron volts The constant determining the absorption of neutrons in boron is obtained from the measurements of the absorption of thermal neutrons described in an earlier section.

The constant α determining the absorption of neutrons in silver is not so easy to obtain. Amaldi and Fermi (1936) have found the value $\mu/\rho = 20$ cm.^{2} gm. $^{-1}$ for the absorption of A group neutrons in silver. The energy of these neutrons (which excite the 22 sec. period radioactivity of silver) is found by absorption measurements to be 2·8 V. Measurements carried out by us seem to indicate a very much smaller absorption coefficient in silver for these neutrons, but this work is probably not entirely comparable with that of Amaldi and Fermi, since in these experiments we used detectors having a mass per unit area of 0·525 gm. The specimen used by Amaldi and Fermi had a mass per unit area of 0·057 gm.

We have made calculations of (5) assuming values of 20 $\text{cm}^2\ \text{gm}^{-1}$ and also of 3 $\text{cm}^2\ \text{gm.}^{-1}$. The actual values of these constants do not modify the general conclusions to be drawn from this analysis. In the calculated curves shown in this paper the value $\mu/\rho = 20\ \text{cm}^2\ \text{gm.}^{-1}$ given by Amaldi and Fermi has been used.

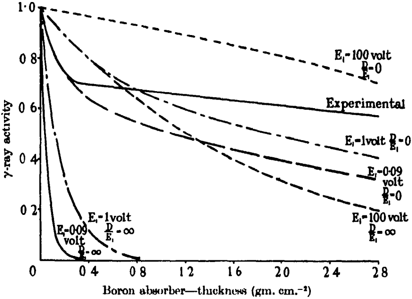


FIG. 5 Comparison of experimental boron absorption curve (using thick silver detector) with theoretical curves calculated for a large number of values of E_1 and D/E_1.

Expression (5) for $A(t)$ has been calculated under these assumptions for a large variety of values of E_1 and D/E_1 for the γ-rays excited in the thick silver detector, and the resulting curves are shown in comparison with the experimental curve in fig. 5. It is evident that under no circumstances can a curve represented by (5) fit the actual experimental curve.

A good fit could be obtained, however, if the magnitude of the slowly varying portion of the curve were not so large in comparison with the activity for zero boron thickness.

It is difficult to see how the assumptions underlying the theory could be modified to provide a reasonable fit to the experimental curve. If, however, the reduced neutron width of the levels Γ_n' were to increase sufficiently with

E_0, a reasonable agreement between the calculated and experimental curves could no doubt be obtained.

It may be, however, that some other process is occurring which gives rise to γ-radiation in the case of collision between fast neutrons and silver atoms. Possibly the most reasonable hypothesis to make is that some of the γ-radiation arises from a type of inelastic scattering of fast neutrons which is not represented in expression (5).* To allow for this it would be necessary to raise the base-line of the experimental curve by an amount corresponding to the intensity of the γ-radiation produced in this way. Applying this hypothesis, a very good fit between the theoretical and experimental curves can be obtained by supposing $E_1 = 0{\cdot}5$ volt, $D/E_1 = 11$, and that 36 % of the total γ-ray activity arises from the independent process. Further, this fit is unique. Thus fig. 6 shows a comparison between the experimental and theoretical curves under these conditions Broadly speaking, the value of E_1 determines the slope of the early part of the theoretical curves, the pro-

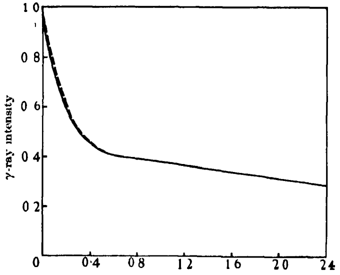


FIG. 6. Comparison of experimental and theoretical curves for γ-rays excited in a thick silver specimen

- - - - Theoretical curve ($E_1 = 0{\cdot}5$ V, $D/E_1 = 11$).

- - - Experimental curve for γ-rays produced by resonance capture assuming 36 % of total observed intensity due to some process other than resonance capture portion of the fast neutron activity due to the postulated independent process determines the slope of the curve for large values of the boron absorber thickness and D/E_1 determines the relative importance of the two terms of equation (5).

---

* Expression (5) takes into account one type of inelastic scattering, viz. that corresponding to capture of the incident neutron followed shortly afterwards by emission of the same or another neutron with lower energy.

In the case of the thin silver specimen the resonance activity arises almost entirely from capture into the lowest energy level, i.e the boron absorption curve due to resonance capture should be almost exponential with a value of the absorption coefficient $\mu/\rho = 7{\cdot}5$ cm.^{2} gm.$^{-1}$ corresponding to a neutron energy of 0·5 V. Superimposed on this exponential, however, will be a small constant component due to γ-rays produced in an inelastic scattering process if the previous hypothesis is valid Assuming the γ-ray intensity due to this process to be proportional to the thickness of the specimen, and that 36 % of the activity produced in the thick specimen arises from this process, the corresponding figure for the thin specimen comes out at 15 % of the total activity. Fig. 7 shows a boron absorption curve calculated on these assumptions for the thin silver detector compared with the experimental points. The agreement is seen to be very reasonable.

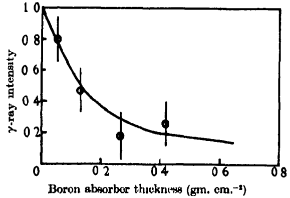


FIG. 7. Comparison of experimental points for thin-silver γ-radiation and curve calculated assuming $E_1 = 0{\cdot}5$ V and 15 % of total γ-radiation to fast neutrons.

No attempt is made in this analysis to distinguish between γ-rays produced in the two silver isotopes. If γ-rays can be produced in both isotopes, the resultant boron absorption curve on the Bethe-Placzek theory would be a combination of two curves of the type of fig. 5. It does not seem that any possible combination of two such curves could give a curve decreasing so slowly for large boron thicknesses as that observed, which indicates the necessity for postulating the production of γ-radiation by some independent process.

The analysis was carried out for silver since it was easiest to determine the curves accurately in this case. Fig. 4, however, shows also curves for the γ-rays excited in arsenic and iodine—elements which consist of a single isotope. From an inspection of the shapes of these curves it is clear that the same difficulties would arise in attempting to obtain a fit to a theoretical curve as arose in the case of silver.

The curves which have been obtained for other elements are similar in form to that for thick silver and can thus be explained in a similar manner. The experimental points are in these cases less accurately determined, however, and so no attempts have been made to assign values for E_1, D/E_1, etc., for them.

It should be pointed out that a constant background of the type of that ascribed to γ-ray production with inelastic neutron scattering could be produced by wholly instrumental means. Thus if neutrons are scattered on to the specimen from directions other than that of the source such a constant background would be produced. On account of the precautions taken to prevent these scattered neutrons from reaching the specimen it is highly unlikely that any γ-ray activity could arise from this cause and certainly nothing of the order of magnitude required to explain the observed effect.

Further, any error in estimating the absorption of γ-radiation from the source in the specimen would produce an uncertainty in the magnitude of the background. It is inconceivable, however, that the uncertainty in this measurement could be sufficiently large to account for that portion of the activity which we have attributed above to inelastic scattering.

ARE THE ENERGY LEVELS WHICH GIVE RISE TO γ-RADIATION IDENTICAL WITH THOSE GIVING β-RADIATION FROM THE SAME NUCLEUS?

Fig. 8 shows the boron absorption curve for neutrons which give rise to γ-radiation in silver together with curves for the absorption of neutrons which give rise to the two β-ray activities (half periods 22 sec. and 138 sec. respectively).

In each case the thickness of the silver detector was 0·5 mm. In the case of the latter two curves precautions similar to those described above for the γ-ray determinations were taken to prevent neutrons scattered from other parts of the room influencing the results. The geometry of the arrangement was reasonably good, the mean obliquity of neutrons passing through the absorber being about 15° to the normal, and only a small correction was necessary to allow for this.*

---

* The obliquity correction in the γ-ray case was also small.

The curves indicate that the absorption in boron of the neutrons producing γ-radiation is greater than that for the neutrons giving rise to the artificial radioactivity, although the measurements do not lie far outside the probable error which, on account of the smallness of the effect, is very large in the case of the γ-ray activity.

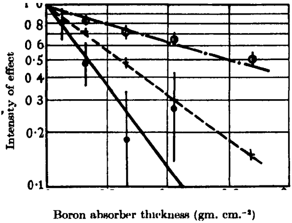


FIG. 8. Comparison of boron absorption curve for the various activities induced in silver by neutrons.

ϕ — -- - γ-ray activity (thin detector)
+-------- 22 sec. β-radioactivity.
ϕ-— - 138 sec. β-radioactivity.

Further light on the question of the identity of the resonance levels which give rise to γ-radiation and those which give rise to the β-radiation has been obtained by measuring the relative activities produced by thermal and non-thermal neutrons in the different cases. The same disposition of paraffin and detector was used in each case, so that the quality of the neutron beam reaching the specimen was the same. The experimental arrangement was similar to that of fig. 1, except that a great deal more paraffin was used round the source to increase the percentage of slow neutrons.

Table III shows the results obtained for this measurement.

TABLE III

Activity	Ratio of activity produced by thermal neutrons to that produced by non-thermal neutrons
22 sec. β-radioactivity	$3{\cdot}8 \pm 0{\cdot}7$
γ-ray emission	$5{\cdot}1 \pm 0{\cdot}9$

The relative activity produced by thermal to that produced by non-thermal neutrons appears to be greater in the case of the γ-ray activity than in the 22 sec. β-ray case, although the two measurements lie just within the limits of probable error. The 138 sec β-activity has not been investigated in this way.

Information on the above question is also obtained from the measurements described in the last paragraph by comparing the number of β-ray electrons and γ-ray quanta produced. For a neutron source consisting of 350 millicurie of Rn + Be a sheet of silver of 0·5 mm thick, 9 × 7 cm. placed close to the counter (enclosed in 2 mm. of lead) produces about 50 impulses per min. The 22 sec β-ray activity measured by bringing the silver specimens after neutron irradiation in approximately the same positions relative to the counter as is the silver during the γ-ray measurements, and extrapolating to zero time after removal from the neutron source is about 80 per min. Allowing for absorption of the β-radiation in the silver and the Al wall of the counter (0·3 mm. of Al), the corrected activity at the counter would come out at about 200 impulses per min. for the particular arrangement of lead and paraffin used.

For γ-ray detection the efficiency of the system (counter + cylindrical lead shield) is estimated at approximately 1 %, based on the approximate relative absorption coefficient of the Ag γ-radiation and its secondaries in Pb. This implies that the number of impulses that would be produced at the counter if the γ-radiation were totally absorbed is 5000 per min , i.e. larger than the number of 22 sec. β-ray electrons by the factor 25.

When equilibrium has been reached the number of 138 sec. β-rays per min. is somewhat less than the number of 22 sec. β-rays. The number of γ-ray quanta produced on neutron capture therefore exceeds the total number of β-rays of both periods by a factor of at least 15, i.e. either fifteen γ-ray quanta are produced per absorbed neutron in addition to a β-ray or the γ-radiation arises from a process different from that which gives rise to the β-radiation.

The first hypothesis does not appear very attractive. In view of this evidence it appears possible that the resonance γ-radiation arises from a process independent of that which produces either of the β-radioactivities, although a conclusive test of this point is scarcely possible until more intense neutron sources are available. Since there are only two known stable isotopes of silver, it is difficult to understand where the two β-radioactivities and the γ-ray effect could arise. It would appear necessary to postulate that the same nucleus may have two sets of energy levels, the capture into one of these giving rise to a γ-radiation, whilst capture into the other produces a β-radioactive nucleus.

A similar phenomenon has been observed by Waring and Chang (1936) in a study of the positron radioactivity induced by the bombardment of $^{27}_{13}Al$ by α-particles. There are two known methods of disintegration of $^{27}_{13}Al$ by α-particles, viz.

$$(1)\ ^{27}_{13}Al + ^{4}_{2}He \rightarrow ^{30}_{14}Si + ^{1}_{1}H,$$

$$(2)\ ^{27}_{13}Al + ^{4}_{2}He \rightarrow ^{30}_{15}P + ^{1}_{0}n.$$

The resonance levels in the case of the first reaction have been known for some time. Waring and Chang investigated the resonances in the second case and found them different from those for the first. Since in each case the intermediate nucleus is the same, they have interpreted this to imply that nuclear resonance phenomena depend not only on the energy of the incident particle but also on the nature of the final state of the system. Our work points to the possibility of a similar conclusion in the case of neutron capture.

We wish to thank Professor T. H. Laby, F.R.S., for his interest and advice during the progress of the work. We are also indebted to Dr H. C. Webster and Mr F. G. Nicholls of the Radio Research Board for help in connexion with the counting circuits.

The radon was supplied by the Commonwealth Radium Laboratory, and our thanks are due to the officers of the laboratory (Messrs T. H. Oddie and W. N. Christiansen) for preparing the sources for this work.

Two of us (R. D. H. and A. A. T.) were holders of Dixson Research Scholarships during the course of this investigation.

SUMMARY

Measurements have been made of the absorption in boron of those neutrons which excite γ-rays in specimens of cadmium, silver (of two thicknesses), arsenic, antimony, iodine and mercury. The measurements were carried out using thermal neutrons in the case of cadmium and silver (thin specimen) and non-thermal neutrons in the case of silver, arsenic, antimony and mercury.

The curves in the latter case were always of the same general type, consisting of an initial component rapidly decreasing with boron absorber thickness superimposed on a component varying very slowly with boron thickness. It is shown that the initial component is to be associated with a nuclear resonance level corresponding to very small neutron energy, while the other component is due to neutrons captured into resonance levels of

greater energy. The curves are interpreted in terms of the Bethe-Placzek formulation of Bohr's theory of nuclei.

In the case of silver, absorption measurements in boron have in addition been made of those neutrons which excite the β-radiations of 22 sec. and 138 sec. half-period.

The question of whether a nucleus which on neutron capture can emit both β- and γ-radiation will have the same resonance levels corresponding to both methods of reorganization is discussed. For silver the evidence points to the existence of separate sets of energy levels for γ-ray emission and β-ray decay, but it is not sufficiently conclusive to settle this point finally.

References

Amaldi, E. and Fermi, E. 1936 *Phys. Rev.* **50**, 899.
Bethe, H. A. and Placzek, G. 1937 *Phys. Rev.* **51**, 450.
Bohr, N. 1936 *Nature, Lond.* **137**, 344.
Breit, G. and Wigner, E. 1936 *Phys. Rev* **49**, 519
Burhop, E H. S, Hill, R D. and Townsend, A A. 1936 *Nature, Lond.*, **138**, 1094.
Danysz, M, Rotblat, J, Wertenstein, L and Zyw, M. 1934 *Nature, Lond.*, **134**, 970.
Ehrenberg, W. 1935 *Nature, Lond*, **136**, 870.
Goldsmith, H. H. and Rasetti, F. 1936 *Phys. Rev.* **50**, 328.
Lea, D E. 1935 *Proc. Roy. Soc.* A, **150**, 637.
Livingston, M. S. and Hoffman, J G. 1937 *Phys. Rev.* **51**, 1021.
Mitchell, D P. 1936 *Phys. Rev.* **49**, 453.
Preiswerk, P and von Halban, H. 1936 *C.R.* **202**, 840.
Rasetti, F, Segrè, E, Fink, G. A., Dunning, J. R. and Pegram, G. B. 1936 *Phys. Rev.* **49**, 104.
Rasetti, F., Mitchell, D. P., Fink, G A. and Pegram, G. B. 1936 *Phys. Rev.* **49**, 777.
Waring, J. R. S. and Chang, W. Y 1936 *Proc. Roy Soc.* A, **157**, 652.

The formation of mercury molecules. II

By F. L. Arnot, Ph.D., *Lecturer in Natural Philosophy* and
Marjorie B. M'Ewen, B.Sc., *The University, St Andrews*

(*Communicated by H. S. Allen, F.R.S.—Received* 17 *December* 1937)

Introduction

In the first paper on this subject Arnot and Milligan (1936), using a mass spectrograph, showed that diatomic mercury molecules were formed by attachment of excited atoms to normal atoms, the products of the collision being an ionized molecule and a free electron It was shown that for this process to occur the excited atom must be in an energy state about 9·5 V above the normal

The primary object of the work described in this paper was to determine this appearance potential more precisely by using the balanced space charge method. The lowest excited state of an atom which, on collision with a normal atom, leads to the formation of a molecular ion is shown to be the $6s8p\ 4\ {}^1\mathrm{P}_1$ state of 9·722 V energy. In addition we have detected molecular ions formed by the attachment of two excited atoms, and evidence is put forward to show that only atoms in P states form attachments with normal atoms which lead to the formation of molecular ions.

Apparatus

A scale drawing of two forms of the apparatus, which gave identical results, together with a wiring diagram, is shown in fig. 1. All metal parts, including gauzes, filament leads and connecting wires, were made of nickel, spot-welded where necessary The cylinders C_1 and C_2 were of identical construction except for the gauze window in C_1. The caps on the ends of these cylinders were pressed out of one piece of nickel to give a perfect fit, and then were spot-welded on. The cylinder C_1 was fitted into a rectangular box folded from a single sheet of nickel so as to prevent any electrons from the filament F_1 reaching the outside of C_2. This box contained the gauze window G_1 between F_1 and the field-free space S where the ions were produced. The cylinder C_2 was supported from C_1 by quartz rods which insulated it from the rest of the apparatus. The filament F_1 was a straight tungsten wire 0·15 mm. in diameter and 12 mm. long. The filaments F_2 and F_3, which were connected in parallel, were each of tungsten 0·1 mm. in diameter and 12 mm. in length They were co-axial with the cylinders C_1 and C_2, and their leads were insulated from the cylinders by short quartz tubes passing through the end-caps.

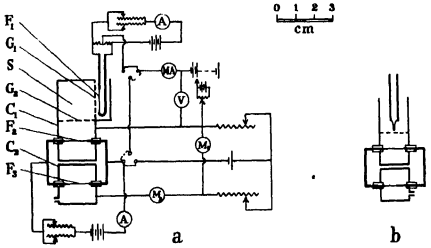


FIG. 1. Apparatus and wiring diagram.

The apparatus was contained in a large pyrex tube fitted with a 5 cm. ground-glass joint, so that the whole apparatus could be withdrawn for adjustment and filament renewal. The joint was water-cooled and a low vapour pressure Apiezon grease was used. The tube was connected to a mercury diffusion pump backed by a Hyvac pump, and to a McLeod gauge.

A pool of mercury was kept at the lower end of the tube which could be heated by a non-inductively wound electric furnace. To prevent rapid diffusion of mercury vapour at high pressures a nickel-plated copper disk, just small enough to pass through the ground joint, was fitted between the apparatus and the outlets to the pumps and McLeod gauge. This arrangement was found to be quite satisfactory for vapour pressures up to about 0·03 mm of Hg at 0° C., but to obtain pressures of the order of 0·1 mm. the following modification was made. The apparatus shown in fig. 1 *a* was slightly reduced in size so that it could be enclosed in a nickel cylinder 3 cm. in diameter and 6 cm. long. This cylinder was closed by a tight-fitting cap at each end, the only opening being a small aperture for the entry of the leads to the filament F_1. The other leads were taken out through quartz tubes fitting tightly into the containing cylinder. When mercury was placed inside this cylinder it was found that the heat radiated from the filaments was sufficient to produce a high vapour pressure.

The vapour pressure corresponding to different temperatures of the furnace was found by using an ionization gauge substituted in place of the apparatus in the pyrex tube. The gauge was of the external collector type and was calibrated against the McLeod with nitrogen. Fig. 2 shows the variation of mercury vapour pressure with furnace temperature. The broken curve represents the saturation vapour pressure corresponding to the same furnace temperatures reduced by a factor of one-quarter, these values being taken from the "International Critical Tables" and reduced to 0° C. The curves show that for high vapour pressures the actual pressure in the tube was approximately one-quarter of the saturated value at the same furnace temperature. These curves and all values of the vapour pressure given in the paper are reduced to 0° C.

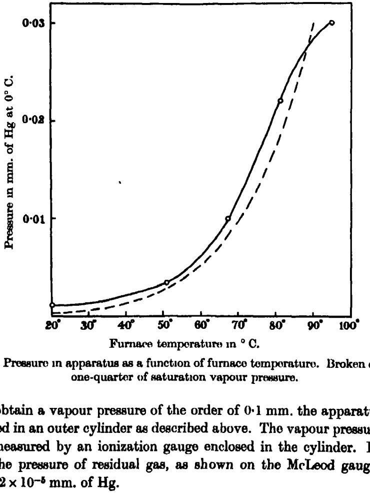


FIG. 2. Pressure in apparatus as a function of furnace temperature. Broken curve is one-quarter of saturation vapour pressure.

To obtain a vapour pressure of the order of 0·1 mm. the apparatus was enclosed in an outer cylinder as described above. The vapour pressure was then measured by an ionization gauge enclosed in the cylinder. For all runs the pressure of residual gas, as shown on the McLeod gauge, was below 2×10^{-5} mm. of Hg.

EXPERIMENTAL PROCEDURE

After the usual flashing and running in of all filaments the current through the filaments F_2 and F_3 was set to give a saturated emission from each filament of about 0·1 mA. The potential between F_2 and F_3 and their respective cylinders was then reduced to 2 V, resulting in the emissions from F_2 and F_3 being strongly space-charge limited to a value of about 0·01 mA. The emissions from F_2 and F_3 were then balanced by the bridge arrangement shown in fig. 1 which incorporated two 10,000 ohm resistance boxes and the galvanometer M_1 having a sensitivity of $7{\cdot}8 \times 10^{-11}$ amp./mm. In this way any fluctuations in the potential of the cells supplying the heating current to F_2 and F_3 were eliminated. The emission from F_3 alone was read on the galvanometer M_2 having a sensitivity of $9{\cdot}7 \times 10^{-9}$ amp /mm. so that its constancy during a run could be checked. Both galvanometers were provided with universal shunts so that their sensitivities could be suitably altered.

Electrons from F_1 were accelerated up to the gauze G_1 by a potential V_0 applied to the centre of the filament as shown in fig. 1. V_0 was read on a Weston standard voltmeter. Some of the ions produced in the field-free space S diffused through the gauze G_2 and, by neutralizing the space charge around F_2, increased its emission which was indicated by the deflexion of the galvanometer M_1. This method is known to be a very sensitive detector of positive ions, since it has been shown that a single positive ion neutralizes the space charge of as many as 10^4 to 10^6 electrons, for the ions revolve in spiral orbits around the filament, making many loops before finally hitting the fine wire

When V_0 was below the ionization potential the emission from F_1 was space-charge limited and consequently increased as V_0 was increased. When the emission from F_1 was kept as constant as possible by readjusting the filament current after each change in V_0, it was found that the number of electrons passing through the gauze G_1 into the field-free space S varied in an erratic manner. This was determined by insulating the gauze G_1 from the rest of the apparatus, and then measuring the currents to G_1 and to the box behind G_1.

When the emission from F_1 was allowed to increase as V_0 was increased it was found that the number of electrons passing through the gauze increased steadily with V_0. This steady increase was linear up to $V_0 = 10{\cdot}4$ V, the ionization potential, after which the rise became progressively less steep as the emission reached its saturation value of about 0·16 mA. It was thus found to be more satisfactory not to attempt to keep the emission from F_1 constant, but to allow it to increase steadily as V_0 was increased.

Results

(1) *Measurements in vacuum*

Measurements in vacuum were made by immersing the tube containing the apparatus in a large Dewar flask containing a mixture of solid carbon dioxide and alcohol. V_0 was increased from 0 to 12 V in steps of 0·5 V, and the change in the emission from F_2 was measured by the galvanometer M_1.

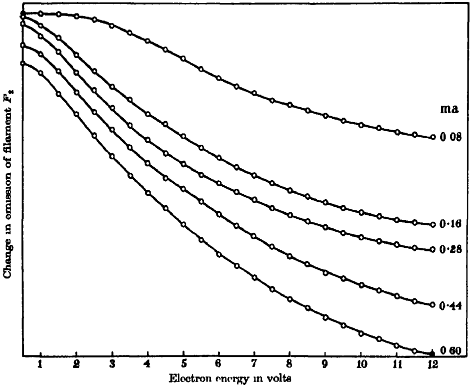


Fig. 3. Results in vacuum for various values of the saturated emission in milliamperes given by filament F_2.

The results obtained are shown in fig. 3. The number on the right of each curve is the saturated emission of F_2 in milliamperes, and thus represents a measure of the temperature of the filament which must be increased in order to increase the saturated emission. The curves have been slightly displaced in a vertical direction so as to separate the points at low values of V_0.

It is seen from fig. 3 that the emission from F_2 falls off steadily as V_0 is increased, and that the rate of decrease increases with the temperature of the filament F_2. It was found that the rate of decrease in the emission from F_2 increased also when the emission from F_1 was increased.

The reason for this behaviour is due to the fact that the emission from F_2 is not space-charge limited over its entire length. The temperature of the filament is highest in the centre, and the emission drawn from this portion is limited by space charge; but the emission from the ends of the filament, which are cooled by the leads, is temperature limited. Electrons from F_1 can have no effect on the space-charge limited emission from the centre of F_2, but they decrease the temperature-limited emission from the ends of F_2 by virtue of their space charge Now as V_0 is increased the number of electrons penetrating the gauze G_1 and so affecting the emission from the ends of F_2 is increased, thereby causing the emission from F_2 to decrease as V_0 is increased The same effect is caused by increasing the emission from F_1

If the temperature gradient at the two points, on either side of the central portion of the filament, at which the emission just ceases to be space-charge limited is small, then the electrons from F_1 will produce more effect than if the temperature gradient is large, since the length of filament giving a temperature-limited emission will be larger the smaller the temperature gradient. The temperature gradient at these two points will be smaller the lower the temperature of the centre of the filament, that is the smaller the saturated emission from the filament. This is borne out by the curves in fig 3 which show that the fall off in the emission from F_2 is less the smaller the saturated emission that could be drawn from the filament.

To minimize this effect the filament current for F_2 and F_3 was kept low in future runs, the saturated emission from each of these filaments being .0·08 mA and the space-charge limited value being 0·01 mA. It will be seen from the uppermost vacuum curve in fig 3 that the fall off in the emission of F_2 beyond $V_0 = 3$ V is not far from linear for this value of the saturated emission. For many of the runs in mercury vapour this fall off in the emission from F_2 as V_0 was increased was found to be strictly linear up to the point at which ions were formed. This enabled us to determine the ionization potential more accurately than would otherwise have been possible by extrapolation of the curve as in fig. 6.

(2) *Measurements in mercury vapour*

In fig. 4 are shown three curves obtained at different pressures of mercury vapour. The curves represent the change in the emission from the filament F_2, measured by the galvanometer M_1, caused by altering the energy V_0 of the electrons in the primary beam from F_1. Curve (*a*) is for a pressure of 0·001 mm., curve (*b*) for 0·031 mm., and curve (*c*) for 0·11 mm. of Hg at 0° C.

Considering the low-pressure curve (*a*) we see that as the energy of the primary electrons V_0 is increased up to 10 V the emission from F_2 falls off steadily as occurred in vacuum. At 10·5 V the curve begins to rise sharply, showing the presence of positive ions which decrease the space charge around F_2. The behaviour of the medium-pressure curve (*b*) is similar except that positive ions now appear when V_0 is approximately 9·5 V. This movement of the point where the curve begins to rise through about 1 V when the pressure was raised by a factor of ten from that prevailing at room temperature occurred in all pairs of curves taken at the two pressures. In addition, a distinct change of slope occurred in the higher-

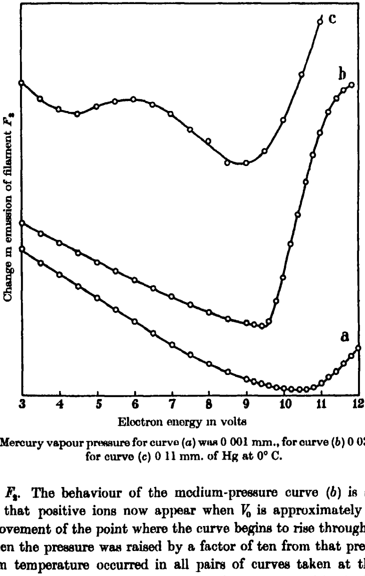


Fig. 4. Mercury vapour pressure for curve (*a*) was 0·001 mm., for curve (*b*) 0·031 mm., for curve (*c*) 0·11 mm. of Hg at 0° C.

pressure curves at the value of V_0 where the lower-pressure curves begin to rise. This change of slope in the higher-pressure curves at a point about

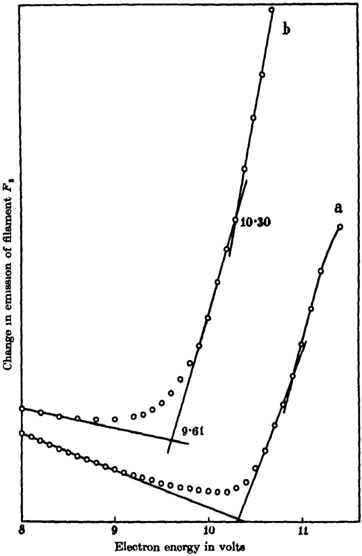


Fig. 5. Showing onset of atomic and molecular ionization. Curve (*a*) obtained with apparatus of fig. 1*a*, curve (*b*) with that of fig. 1*b*.

1 V above the bend in the curves is shown in two typical curves in fig. 5. The change of slope in the rising portion of the curve did not occur in any of the curves taken at room temperature.

It has been shown by Arnot and Milligan (1936) that mercury molecular ions are formed by a collision between a normal atom and an excited atom. Since this collision must occur within the lifetime of the excited atom, about 10^{-8} sec., the probability of molecular ions being formed at low pressure is extremely small. We therefore identify the point at which the low-pressure curves rise as the ionization potential of the mercury atom. Since this point is replaced by a change of slope in the rising portion of the high-pressure curves we identify this inflexion as the ionization potential of the atom. The point at which the high-pressure curves begin to rise is then the appearance potential of the molecular ion, which Arnot and Milligan showed by magnetic analysis to have a value of approximately 9·5 V.

We can now obtain the appearance potential of the molecular ion from the high-pressure curves by subtracting the difference in energy between the point at which the curves begin to rise and the point above this where the curves change slope from the known ionization potential of the atom, 10·39 V. The value obtained in this way is independent of any contact potentials existing in the apparatus. Seven values of the appearance potential obtained in this way from different curves by means of the two forms of the apparatus shown in fig 1 are given in Table I together with the pressure in mm. of Hg at 0° C. at which the curves were taken

TABLE I

Apparatus of fig. 1 *a*		Apparatus of fig. 1 *b*	
Pressure mm. of Hg at 0° C.	Appearance potential volts	Pressure mm. of Hg at 0° C.	Appearance potential volts
0·026	9·71*	0·030	9·70†
0·020	9·62	0·027	9·64
0·017	9·77	0·020	9·65
0·009	9·68		

* From curve (*a*) of fig. 5 † From curve (*b*) of fig. 5.

Three other curves obtained with a slightly modified form of the apparatus shown in fig. 1 *a* gave values of 9·78, 9·82, 9·64. The mean of these ten results is $9{\cdot}70 \pm 0{\cdot}01$.

It has been shown by Arnot and Milligan (1936) that the molecular ion is formed by the attachment of an excited atom to a normal atom. The process may be written symbolically as

$$\mathrm{Hg}' + \mathrm{Hg} \rightarrow \mathrm{Hg}_2^+ + \epsilon, \qquad (1)$$

in which Hg′ denotes an excited atom. The appearance potential represents the minimum energy that the excited atom must possess for the above process to occur.

It was found by Rouse and Giddings (1926), whose results were checked and developed by Houtermans (1927) and by Foote (1927), that ions are produced in mercury vapour when irradiated by its resonance radiation 2537 A, of which the photons have only 4·86 V energy. The ionization was found to be proportional to the square of the intensity of the light. To explain these results Houtermans suggested that an excited atom in the $2\,^3P_1$ state of 4·86 V energy may combine with a metastable atom in the $2\,^3P_0$ state of 4·66 V energy to form an ionized molecule and a free electron. He obtained evidence to show that a collision between two atoms both of which are in the metastable state $2\,^3P_0$ does not lead to ionization. The effective process may be written as

$$\mathrm{Hg}'\,(2\,^3P_1) + \mathrm{Hg}'\,(2\,^3P_0) \rightarrow \mathrm{Hg}_2^+ + \epsilon. \qquad (2)$$

Since this process involves the collision of *two* excited atoms we should expect it to occur at a considerably higher pressure than that necessary for the occurrence of process (1) which involves only one excited atom.

On raising the mercury vapour pressure in our apparatus to 0·11 mm. we obtained curves showing that ions were produced by electrons of energy between 4 and 5 V. A typical curve obtained at this high pressure is shown as curve (*c*) in fig. 4. By subtracting a vacuum run from this curve we obtain the corrected form shown in fig. 6 as curve (*a*). Since the rise in the curve of fig. 4 between 4 and 5 V cannot set in below the resonance potential of 4·86 V a correction to the voltage scale of fig 6 of + 0·65 V has been made This correction, which can be accounted for by contact potentials, makes the steep rise between 9 and 10 V set in at 9·7 V, the mean value obtained from the results given in Table I. The broken curve in fig. 6 represents the theoretical excitation function of the $2\,^3P_1$ state obtained by Penney (1932)

The close agreement between this excitation function and the hump in the curve of fig. 6, together with the fact that this hump only appears when the pressure is raised to about 0·1 mm., leaves little doubt that the molecular ions produced by electrons of energy between 4·86 and 9·70 V are formed by collision between a $2\,^3P_1$ atom and a metastable $2\,^3P_0$ atom as suggested by Houtermans to account for the ionization produced in mercury vapour when irradiated by its resonance radiation.

DISCUSSION

During the progress of the present work a paper appeared by Snavely (1937) on the same subject using an apparatus essentially similar to that shown in fig. 1 *b*. His published curve taken at a pressure of 0·09 mm. is reproduced as curve (*c*) in fig. 6 Comparing this with our curve (*a*) taken at approximately the same pressure, 0·11 mm , we see that the two curves are of approximately the same shape, but that Snavely's curve is moved bodily to the right relative to our curve

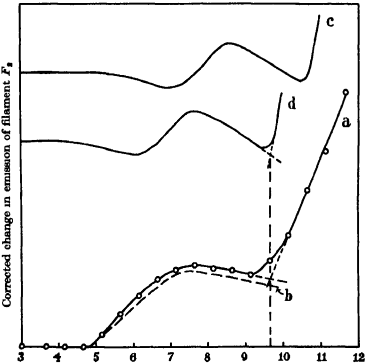


FIG. 6 (*a*) Curve (*c*) of fig. 4 corrected by subtraction of a curve taken in vacuum. (*b*) Penney's theoretical excitation function for the $2\,^3P_1$ state. (*c*) Snavely's original curve. (*d*) Snavely's curve with suggested voltage scale correction.

Snavely corrected the voltage scale for contact potentials by moving his curve until the second minimum appeared at 10·39 V, for he interpreted the steep upward rise after this minimum to indicate primary ionization of the atom. He remarked that an inflexion occurred in this steep upward rise at 11·2 V when his curve was so adjusted but offered no explanation of this. Our results indicate that this inflexion, which we show in fig. 5, represents the onset of direct atomic ionization, and therefore should be set to appear in Snavely's curve at 10·39 V. When the voltage scale of Snavely's curve is readjusted to bring this inflexion down from 11·2 to 10·39 V the beginning of the steep upward rise after the second minimum occurs at 9·7 V, which agrees with our value of 9·70 V for the onset of molecular ionization by process (1), p. 141.

Snavely's curve adjusted in the above way is shown as curve (*d*) in fig. 6. The maximum of his peak now occurs at precisely the same value of the electron energy, 7·65 V, as does the maximum of our peak in curve (*a*). However, this peak in Snavely's curve appears to set in at 6·1 V, whereas we find it to begin at 4·9 V. Since Snavely reports no vacuum runs it appears likely that the fall* in his curve (*d*) between 4·1 and 6·1 V would be steeper in vacuum, so that the peak might really set in at the same value as we find. Our results indicate that this peak is due to molecular ionization by process (2), p. 142.

For the molecular ions formed by process (1) Arnot and Milligan (1936) obtained a curve showing the probability of formation of these ions as a function of the electron energy. This curve contained two maxima, a sharp one at 11·5 V and a broad one at 47 V. Since the appearance potential of the molecular ion formed by this process is 9·70 V the excited atom must possess 9·70 V energy before it can combine by process (1) to form an ionized molecule. The excited state of the mercury atom having an energy nearest to this amount is the $6s8p\ 4\,^1P_1$ state which has 9·722 V energy. Just above this singlet state there is the triplet state $6s9p\ 5\,^3P_{0,1,2}$ of which the energies are 9·792, 9·796, 9·817 V (Bacher and Goudsmit, 1932).

Schaffernicht (1930) gives an optical excitation function for the singlet 4 P state of 9·722 V energy which has a broad maximum at 45 V, which would account for the broad maximum at 47 V in Arnot and Milligan's curve. No excitation functions are given for the triplet 5 P states, but, since these states are excited by electron exchange, the excitation functions will rise to a sharp maximum just above their critical potentials, and would therefore account for the sharp maximum at 11·5 V in the efficiency of molecular ionization curve of Arnot and Milligan, and also for the marked falling off of our ionization curve (*b*) in fig. 4 at 11·5 V. Above 10·39 V this curve represents the efficiency of molecular plus atomic ionization, and consequently only an inflexion, not a sharp maximum, is to be expected at 11·5 V. This inflexion does not appear in the curves taken at low pressure where the molecular ionization is negligible.

We thus reach the conclusion that the molecular ions discovered by Arnot and Milligan when mercury vapour is bombarded by electrons of energy between 9·70 and 200 V can be entirely accounted for by the two following processes:

---

* This fall, which appears also in our curves, has been shown to be due to the emission from the ends of the filament not being space-charge limited.

$$\mathrm{Hg}'\ (4\ {}^1\mathrm{P}_1) + \mathrm{Hg} \rightarrow \mathrm{Hg}_2^+ + \epsilon, \tag{a}$$

$$\mathrm{Hg}'\ (5\ {}^3\mathrm{P}) + \mathrm{Hg} \rightarrow \mathrm{Hg}_2^+ + \epsilon. \tag{b}$$

Process (*a*) accounts for the maximum in Arnot and Milligan's efficiency of molecular ionization curve at 47 V and process (*b*) for the sharp maximum at 11·5 V. We cannot say for certain from our experimental results that an excited atom in an S, D or F state above the 5 ^{3}P state does not lead by collision with a normal atom to an ionized molecule, but we can say that the two processes (*a*) and (*b*) involving only the 4 ^{1}P state and the 5 ^{3}P states are sufficient to account fully for the results of Arnot and Milligan. It will be shown below that there is some evidence that an excited atom in an S, D or F state does not form an ionized molecule

We can now fix the value of the appearance potential, which we have found in this paper to be 9·70 ± 0·01 V, accurately from process (*a*). The appearance potential is the energy of the 6*s*8*p* 4 ^{1}P$_1$ state, namely 9·722 V, since this is the state nearest in energy to the experimentally determined value

The results given in fig 6 lead to the conclusion that ionization is also produced by the process

$$\mathrm{Hg}'\ (2\ {}^3\mathrm{P}_1) + \mathrm{Hg}'\ (2\ {}^3\mathrm{P}_0) \rightarrow \mathrm{Hg}_2^+ + \epsilon. \tag{c}$$

The results of Arnot and Milligan show that the ions produced by processes (*a*) and (*b*) are definitely molecular, since they were detected by a mass spectrograph, but the ions produced by process (*c*) have not been analysed by this means. It might therefore be suggested that the collision process represented by the left-hand side of (*c*) leads to atomic ions. The process

$$\mathrm{Hg}'\ (2\ {}^3\mathrm{P}_1) + \mathrm{Hg}'\ (2\ {}^3\mathrm{P}_0) \rightarrow \mathrm{Hg}^+ + \mathrm{Hg} + \epsilon$$

is energetically impossible since the combined energy of excitation is only (4·86 + 4·66) = 9·52 V, which is 0·87 V short of the ionization potential of the atom. However, the process

$$\mathrm{Hg}'\ (2\ {}^3\mathrm{P}_1) + \mathrm{Hg}'\ (2\ {}^3\mathrm{P}_0) \rightarrow \mathrm{Hg}^+ + \mathrm{Hg}^-$$

is energetically possible, since the electron affinity of Hg is given by Glockler (1934) as 1·79 V. The total energy of the right-hand side is now (10·39 − 1·79) = 8·60 V, which is less than the energy of the left-hand side.

The excess energy of excitation to be carried away as kinetic energy by the products of the process is $(9{\cdot}52 - 8{\cdot}60) = 0{\cdot}92$ V

Theoretical investigations of electron transfer such as is involved in this process, which have been made by Massey and Smith,* show that electron transfer does not become at all probable until the kinetic energy with which the particles collide is greater than ten times the excess excitation energy which must be carried away by the particles after the collision. Now in the above process the kinetic energy with which the particles collide is only their thermal energy which is about 0·1 V, and the excess excitation energy to be carried away by the separating particles is 0·92 V, that is ten times as great instead of one-tenth as great In view of this we consider that the ions are formed in the molecular state by process (*c*).

From process (*c*) we see that the ionization potential of the mercury molecule must be equal to, or less than, $(4{\cdot}86 + 4{\cdot}66 + D) = 9{\cdot}52 + D$ volts, where D is the work of dissociation of the normal molecule. Since Houtermans has shown that the collision of two $2\,^3P_0$ atoms does not lead to ionization the ionization potential must be greater than $9{\cdot}32 + D$ volts

A single atom excited to a state of 9·52 V has thus sufficient energy to form an ionized molecule when it collides with a normal atom. We have shown, however, that no ions are formed by collision with normal atoms until the excited atom has 9·70 V energy. The *probable* error in our value is only 0·01 V. We see from the values given in Table I that the *maximum* error could not reduce this value to 9·52 V. Now the excited states of a mercury atom lying within the energy range from 9·52 to 9·722 V are all S, D or F states. The first P state with an energy greater than 9·52 V is the $6s8p\ 4\,^1P_1$ state of energy 9·722 V. Therefore, although the excited atom in one of the S, D or F states between 9·52 and 9·722 V has sufficient energy to form an ionized molecule, no ion is formed until the atom reaches a P state Since the processes (*a*), (*b*) and (*c*) which all involve P atoms are sufficient to account for all the molecular ions observed in mercury vapour the conclusion seems inevitable that the excited atom or atoms must be in a P state for attachment to occur with the production of a molecular ion.

In conclusion we should like to express our thanks to Professor H. S. Allen who kindly made the necessary arrangements to enable one of us (M. B. M'E.) to work in the Natural Philosophy Department.

---

* We are indebted to Dr R. A. Smith for information on this subject.

SUMMARY

An investigation of the formation of ionized mercury molecules has been made by the balanced space-charge method. The results obtained in this work, together with those previously obtained with a mass spectrograph by Arnot and Milligan, show that molecular ions are formed by the three processes

$$\mathrm{Hg}'\,(2\,^3\mathrm{P}_1) + \mathrm{Hg}'\,(2\,^3\mathrm{P}_0) \rightarrow \mathrm{Hg}_2^+ + \epsilon, \tag{1}$$

$$\mathrm{Hg}'\,(4\,^1\mathrm{P}_1) + \mathrm{Hg} \rightarrow \mathrm{Hg}_2^+ + \epsilon, \tag{2}$$

$$\mathrm{Hg}'\,(5\,^3\mathrm{P}) + \mathrm{Hg} \rightarrow \mathrm{Hg}_2^+ + \epsilon. \tag{3}$$

Ions formed by each of these processes have been detected. The appearance potentials for these three processes are respectively 4·86 and 9·722 V for (1) and (2) and 9·792, 9·796 and 9·817 V for the triplet process (3) Ions formed by process (1) which requires a collision between two excited atoms were detected at a vapour pressure of 0·1 mm of Hg at 0° C Ions formed by processes (2) and (3) which involve a collision between one excited atom and a normal atom were detected at pressures of the order of 0·01 mm.

The ionization potential of the mercury molecule lies between $9{\cdot}52 + D$ and $9{\cdot}32 + D$, where D is the work of dissociation of the normal molecule. Using Winans' (1931) value for D the limits are 9·67 and 9·47 V.

Excited atoms in states other than P states do not apparently form ionized molecules by attachment even though they have more than sufficient energy to do so.

REFERENCES

Arnot and Milligan 1936 *Proc. Roy. Soc.* A, **153**, 359.
Bacher and Goudsmit 1932 "Atomic Energy States" (McGraw-Hill Book Co.), p. 228.
Foote 1927 *Phys. Rev.* **29**, 609.
Glockler 1934 *Phys. Rev.* **46**, 111.
Houtermans 1927 *Z. Phys.* **41**, 619.
Penney 1932 *Phys Rev.* **39**, 467.
Rouse and Giddings 1926 *Proc Nat. Acad. Sci., Wash.*, **11**, 514; **12**, 447.
Schaffernicht 1930 *Z. Phys.* **62**, 106.
Snavely 1937 *Phys. Rev.* **52**, 174.
Winans 1931 *Phys. Rev.* **37**, 897.

Oscillography of adsorption phenomena
III. Rates of deposition of oxygen upon tungsten

By M. C. Johnson and A. F. Henson
Physics Department, University of Birmingham

(*Communicated by M. L. E. Oliphant, F.R.S.—Received* 18 *December* 1937)

1. Introduction

The only temperature coefficient hitherto measured for the atomic layer of oxygen on tungsten is that which demonstrates the 150,000 cal./mol. expended in its desorption. If, however, some energy is required before the difference between this heat of adsorption and the heat of dissociation can be liberated, a further temperature coefficient will exist, giving variability to the "condensation coefficient" or fraction of impinging molecules which become enabled to contribute to this layer. Any such variability extends the notion of "activated adsorption" from the slow processes with which it has become associated to the almost instantaneous reaction of "chemisorption". Since filling of the atomic layer is generally supposed to precede oxidation, and oxidation itself has already been studied through the subsequent volatilization of oxide (Langmuir 1915; Rideal and Wansbrough-Jones 1929), an experimental determination of whether the initial deposition also needs activating would be highly desirable. This would, however, be inaccessible to most technique because of the high speed of recording needed. The oscillographic study of reaction velocity, developed by one of us in collaboration with Dr F. A. Vick, and hitherto applied only to evaporation (Johnson and Vick 1935), offers the opportunity by suitable modifications. It fulfills the following requirements. (1) Reaction time may be kept below 1 sec., avoiding the usual slow contaminations due to gases which strike the walls. (2) The surface may be proved to be clean, within a second of gas being admitted, by observation of its thermionic emission, thus escaping the common ascription of activations to impurity. (3) The gas-covered fraction may be kept small enough to avoid interaction between surface particles. (4) No other substance beside oxygen and tungsten is introduced as reagent or as indicator.

2. EXPERIMENTAL

It will be recalled that thermionic indicators of surface reaction depend upon the properties by which an adsorbed gas layer alters the electron emission of a metal. Our previous experiments consisted essentially of oxygenating a tungsten filament, rigorously evacuating, and then switching from a temperature at which the layer is stable to temperatures at which evaporation is rapid. The saturated emission current from the surface passed through high resistances, the changing potential across the latter being recorded by a cathode ray oscillograph. The oscillograms recording the recovery of emission were transformable into curves of isothermal evaporation of oxygen at each temperature selected, yielding the heat of evaporation as temperature coefficient of the rate of change of gas-covered fraction of the surface. The present experiments invert this procedure, photographing depositions instead, so that the high vacuum has to be replaced by controlled introductions of oxygen Preliminary experiments on the magnetic lifting of a filament in and out of a steady molecular beam were carried out with a spring-mounted electrode system made by Dr F. A. Vick, but the other alternative of attempting a repeatable sequence of pressure variation at a fixed surface was finally adopted. The molecular ray technique was retained only as means of producing suitable fractions of a reservoir pressure by evacuating between two slits which admit the gas to the reaction chamber.

The reservoir *A* (fig. 1), of large capacity, is maintained at various oxygen pressures between 10^{-1} and 10^{-2} mm measured on a McLeod gauge. The piston valve *B* was made and ground for us by Mr W. A. Holland, and when very lightly lubricated with Apiezon "*M*" and held down by a 2 lb spring it is accurate enough to keep for several days a pressure of 10^{-1} mm isolated from the highest vacuum on the inner side. When magnetically actuated the piston opens the large reservoir very suddenly to the smaller reservoir *C* behind the first slit. Slits *M*, *N*, 0·15 mm. wide, are separated by a large central chamber directly evacuated by its own vapour pump. The filament of Thoria-free tungsten with its anode and guard rings in the reaction chamber *D*, is aligned on its flat joint *E* by observation through the collimating window *F* when incandescent. Precautions for purity were taken, including pre-treatment of all metals in hydrogen, baking of *D* and its trap, high frequency degassing, freezing of *E*, etc.

By this apparatus, since scattering is inevitable in an incompletely condensable gas we utilize it in the restricted space between slit and filament to build up transient pressures for *D*. Each of these pressures reaches its

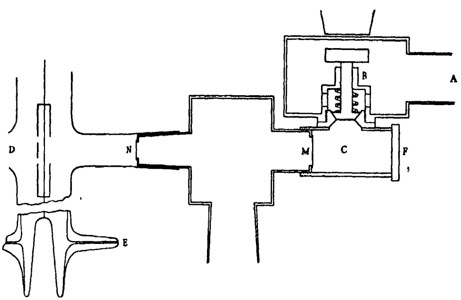
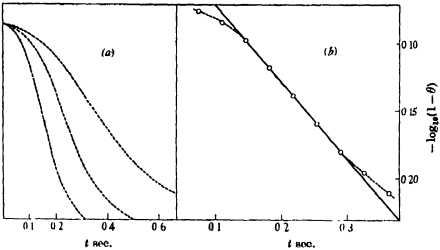


Fig. 1. Apparatus.



(a) (b)

Fig. 2. Adsorption diagrams.

maximum in little more than a tenth of a second after lift of the piston and then remains nearly constant for a short time. This appears in fig. 2*b* when the logarithm of the "clean" fraction of the surface begins to decrease linearly with time. It is seen that linearity lasts less than a second. After this time the fall in pressure, and possibly failure of the law by which fractions of a dilute layer are computed, make any extension to dense covering impracticable. For a long series of observations it is important that the transient nature of these pressures is able to minimize any thinning of the filament; evacuation sets in before the layer reaches semi-saturation and oxidation only becomes appreciable after some weeks. We gain the additional advantages (*a*) that search for laws of deposition is not complicated by temperature coefficients of surface migration, as when the front of the target only is exposed to a uni-directional beam of rays, (*b*) that Magnetron action and distortion of electric field, discovered in our early experiments on moving an emitting filament into a beam, are avoided, so that the electrical changes photographed can be safely ascribed to adsorption if temperature is known to be constant.

The transient pressures in D amount to small multiples of 10^{-4} mm., converting the "instantaneous chemi-sorption" into reactions traceable with the oscillograph A large number of plates were measured for different reservoir pressures and filament temperatures. These yielded curves of emission current against time (such as fig. 2*a*), drawn out to a common scale, a tuning-fork trace being recorded on each plate and checking any changes in calibration. Any small initial value of θ, the fraction of surface gas-covered, was found by changing over to a galvanometer circuit and taking the data for a "Richardson plot" just before each experiment; this was then compared with a standard plot obtained after the filament had undergone its preliminary "ageing".

The curves were found to depend sensitively upon reservoir pressure, as was to be expected; but superposed upon this was found a slighter dependence upon temperature, which was important as proving the existence of activation energy Since changes in pressure affected the rate of adsorption more than considerable changes in temperature, the final estimate of temperature coefficient was reached by finding the empirical law connecting reservoir pressure with deposition rate, and then repeating the whole investigation at a sensibly different temperature, finally checking the first set.

We proceed to show how the quantities i, measured as oscillogram ordinates proportional to emission, can be transformed into the quantities θ, temperature dependence of whose rate of change is a measure of heat of activation.

3 METHOD OF CALCULATION

Let θ be the fraction of the tungsten surface covered with a layer of adsorbed gas Let N be the number of adsorbed particles per cm.^{2} which would fill a "complete" layer, defined e.g. by contact packing or by one-to-one or other relation to lattice points Then θN is the actual number per cm.^{2} and is equivalent to θ/σ, where σ is the "occupied" area for one particle. The status of thermionic approximations to θ is reviewed in a recent paper (Johnson and Vick 1937*b*); subject to the limitations there discussed, we adopt the "work-function" approximation as suitable for values between 0·1 and 0·3, a range which covers our experiments.

In the paper on evaporations it was shown that, if θ is not large, desorption into a vacuum can be represented empirically by expressions which may be summarized here in the form

(*a*) at constant temperature,

$$\log \theta = -t/C;$$

(*b*) at different temperatures,

$$\log C = c'/T.$$

C is then the average duration of an adsorbed state.

These empirical relationships are in accord with equations obtained by assuming that no secondary reactions occur, so that

$$N\frac{d\theta}{dt} = -\theta N \frac{1}{\tau}, \tag{1}$$

$$\frac{d\theta}{dt} = -\theta \frac{1}{\tau_0} e^{-E/kT}, \tag{2}$$

$$\log_e \frac{\theta_t}{\theta_0} = -(t-t_0)\frac{1}{\tau_0} e^{-E/kT}. \tag{3}$$

τ_0 was obtained experimentally as 8×10^{-14} sec. and E as $147{,}000 \pm 3000$ when $k = 2$, i e in cal./mol.

θ was obtained in the usual manner as

$$\theta = \frac{\log i_\theta - \log i_0}{\log i_1 - \log i_0}, \tag{4}$$

depending on its definition, $\phi_\theta - \phi_0/\phi_1 - \phi_0$, and on assumptions as to A in the Richardson equation $i = AT^2 e^{-\phi/kT}$. These assumptions are discussed in the papers referred to (1937). It happens that a knowledge of i_1, the emission from a fully covered surface, was not needed in the study of evaporation, as the required data were sequences of θ at constant temperature and different times, in any of which sequences

$$\frac{\theta_1}{\theta_2} = \frac{\log i_{\theta(1)} - \log i_0}{\log i_{\theta(2)} - \log i_0} \tag{5}$$

The corresponding equations for deposition can be written by enumerating the appropriate factors,

$$N\frac{d\theta}{dt} = (1-\theta)\,\nu K e^{-E'/kT}. \tag{1'}$$

ν is the number of impacts per cm.^{2}/sec from the gas phase. When the system reaches equilibrium, the isotherm formed by adding terms of recombination or evaporation or oxidation may involve $\frac{1}{2}$ instead of unity as the power of ν, according to the nature of such processes which for our small values of θ we neglect. Again, if certain mechanisms are assumed for the dissociation, the power of $(1-\theta)$ might become 2 instead of unity, with the effect of multiplying the right-hand side of equation (3′) by a factor ranging between 0·9 and 0·7. This might extend the range over which our graphs are linear, but has no appreciable effect on our results; we prefer to leave the equation of deposition in its simpler form as we do not wish at this stage to assume any particular mechanism of dissociation.

E' and K express the possibility that not all molecules striking the bare metal may be able to contribute to growth of the layer; they represent the temperature dependence and any other factor in a "condensation coefficient" about which we seek information.

Using the equivalence of N and $1/\sigma$, since the former does not disappear from the deposition equations as it did in evaporation,

$$\frac{d\theta}{dt} = (1-\theta)\,\nu\sigma K e^{-E'/kT}, \tag{2'}$$

and since

$$\frac{d\theta}{dt} = -\frac{d(1-\theta)}{dt},$$

$$\log_e \frac{1-\theta_t}{1-\theta_0} = -(t-t_0)\,(\nu\sigma K e^{-E'/kT}). \tag{3'}$$

Comparing (1′), (2′), (3′), with (1), (2), (3), it becomes apparent that the convenient elimination of i_1 is no longer valid, since the ratio of values of θ, using (4), (5), does not give a ratio of values of $(1-\theta)$. Instead of obtaining, as in evaporation, a value of τ at each T by plotting $\log_e \theta_1/\theta_2$

against time in terms solely of *observed* currents, our observed data now have to be related to an origin only indirectly accessible. The expression corresponding to (5) becomes

$$\frac{1-\theta_1}{1-\theta_2}=\frac{1-\dfrac{\log i_{\theta(1)}-\log i_0}{\log i_1-\log i_0}}{1-\dfrac{\log i_{\theta(2)}-\log i_0}{\log i_1-\log i_0}}. \tag{5'}$$

Thus for the experiments on deposition $\theta=\theta_0$ at $t=0$, where θ_0 is zero or a small known quantity, instead of θ tending to zero at $t=\infty$ as in evaporation; hence the present work requires to be plotted against t the function

$$\log_e\left(1-\frac{\log_{10} i-\log_{10} i_0}{\log_{10} i_1-\log_{10} i_0}\right). \tag{6}$$

Such a plot will yield as slope the quantity $\nu\sigma K e^{-E/kT}$, which formally corresponds to $(1/\tau_0)\,e^{-E/kT}$ in the previous problem.

It is now known that i_1, which we need in the present problem, cannot be measured directly. Since any apparent Richardson plot for the composite surface is distorted by the continuous transition from one adsorptive equilibrium to another, it fails to exhibit the true law of emission for the completely covered metal, which must be reformulated to accord with the contact potential between *OW* and *W*. It has been shown however (Johnson and Vick 1937*a*) that measurements of the temperature dependence of steady emission from *W* in the presence of oxygen are indeed consistent with a work function deduced from Reimann's contact potentials if a reasonable assumption as to the emission coefficients is made. On that basis we have shown how the angular and vertical displacement of the Richardson plot for *OW* relative to that for *W* may be estimated, even though the emission currents which it represents cannot actually be observed.

From such a pair of lines for *OW* and *W* we read off the value of $\log_{10} i_1-\log_{10} i_0$ at each temperature required for the present work. The numerical range is from $-3{\cdot}8$ to $-4{\cdot}2$, and admissible variations in the quantities from which it is derived, e.g. the constant A, will exert only a minor effect upon our results.

4. RESULTS AND CONCLUSIONS

When the experimental data are plotted as described in section 2, the passage from a rising to an approximately constant gas pressure is at once suggested by the form of inflexion leading into exponential curvature (fig. 2*a*). When the logarithm of the decrease of bare surface is computed from these curves by means of equation (6) the linearity of its central portion confirms this suggestion. On our theory a definite slope is controlled by ν in equation (3′) at any given temperature. Fig 2*b* illustrates this feature and also the subsequent divergence from linearity terminating the exponential portion. The divergence is ascribable to the fall of pressure after its maximum, but at $\theta > 0{\cdot}3$ it may also include breakdown in the linear relation between θ and ϕ, the precise limits of which are unknown At large values of θ evaporation and interaction between adsorbed atoms would also cease to be negligible.

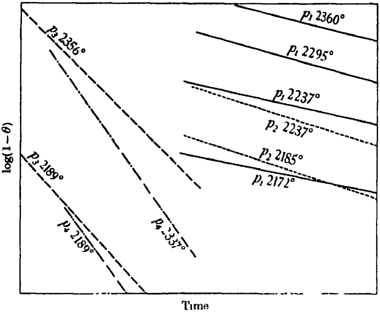


Fig. 3. Variables controlling rate of deposition.

The primary dependence of the rate of adsorption upon pressure rather than upon temperature is shown in the examples of fig. 3, where only the linear portions are reproduced, corresponding each to some constant pressure, to a first approximation each set of such lines at a common pressure has a common slope differing only from another set at any other pressure. For the second approximation, even with the improved methods of section 2, there has to be recognized a slight scattering about any such "common" pressure, tending to mask any effect of temperature; but by selecting pairs of lines whose gas conditions were known from their place in a sequence of admissions to be closely identical, and plotting the logarithm of their rate of deposition on a large scale against I/T (fig. 4*a*), temperature differences all of one sign became evident, two sets for instance yielding heats of activation of 18,000 and 22,000 cal./mol. Since the usual method of plotting against reciprocal temperature suffers seriously in this case from small fluctuations in the gas conditions, these rough estimates were not regarded as more than establishing the existence of an activation, and were supplemented by the following. Two temperatures were selected, after preliminary

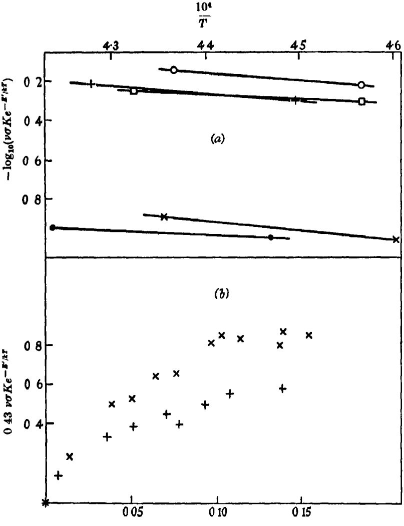


FIG. 4. Temperature dependence. In fig. 4 (*b*) × indicates $T=2272°$, + indicates $T=2122°$.

calculations based on desorption rates, to be widely separated without allowing evaporation or other secondary effect to become appreciably disturbing, and new sets of oscillograms were taken at a number of suitable pressures for each temperature. The rates per second of $\log(1-\theta)$, giving $\nu\sigma K e^{-E'/kT}$, were then plotted against reservoir pressure for each temperature. The two curves (fig. 4*b*) obviously are empirical representations of the fact that the more copious gas admissions are the more rapidly evacuated, and are therefore non-linear, but the laws of pumping will not alter from one filament temperature to the other, especially as the first curve was repeated after the second. Hence each curve may be taken as smoothing the unavoidable fluctuations in the dominant variable or gas pressure, allowing the separation between the two curves to demonstrate the dependence upon the secondary variable or temperature Ordinates were compared at different places along these graphs, yielding a mean heat of activation of 24,500 cal./mol. Repetitions suggest a wide latitude ± 3000 which is not merely due to the low accuracy of extracting the temperature dependence from the pressure dependence; in the earlier work on evaporation we had found that the heat of desorption could rise from the 150,000 of an "aged" but not eroded wire to nearly 200,000 in its later life, and in the present case there is some indication that probability of condensation alters with ageing. It is not unlikely to vary so much with micro-crystalline structure that precision in any particular determination would be misleading, and attempts to distinguish intermediate steps of the temperature range could add little.

The considerable differences in initial θ for sensibly similar lines, e.g the low pressure set on the right of fig 3, indicate that up to 0·3 the progress of condensation is independent of the presence of already adsorbed material. Rates of deposition so nearly alike when initial θ for one experiment exceeds final θ for another, at pressures too low for temperature to be the dominant variable, lend some support to two of the most inaccessible assumptions in thermionic studies of surface reaction (*a*) that in the main a single process is being traced, not a transition to interaction between surface particles, (*b*) that the thermionic approximation to θ, namely $\phi-\phi_0/\phi_1-\phi_0$, is not seriously vitiated in our range by distortions such as are well known in the case of alkalis at large values of θ.

It is useful to estimate the order of magnitude of the number of gas impacts at the filament itself, as indicating the relative importance of K and E' or of temperature-dependent and independent terms in the probability of condensation. From our graphs, when $\nu\sigma K e^{-E'/kT}$ is nearly 2, the fraction $e^{-E'/kT}$ is 5×10^{-3}. σ is of the order 10^{-15}. Hence ν is about 3×10^{17} per cm.^{2}/sec., if K is unity. If K is less, ν is correspondingly greater. 10^{17}

implies a pressure of about 3×10^{-4} mm., which is a reasonable fraction of our measured reservoir pressures considering the size of slits, etc. in use and the time of building up of pressure. The rate of thinning of filament, compared with that resulting from higher pressures showing visible discharges, indicates that actual pressures at the filament cannot much exceed this estimated amount; hence K need not be much less than unity. The temperature dependent term, which is between 10^{-2} and 10^{-3} in the vicinity of 2200° abs., may then be regarded as the main reason why the condensation coefficient as a whole is small.

It has been assumed throughout that the temperature of the surface is completely controlled by the filament heating current, whose constancy was maintained with precision at each selected value This assumption would fail if any of the following contributed appreciably to the thermal state of the filament: (*a*) liberation of heat of adsorption, (*b*) cooling by thermal conduction in the gas, (*c*) cessation of the cooling normally caused by electron emission, when that emission is inhibited by the oxygenation. By considering the magnitudes of θ, the expenditure of energy which maintains the thermal equilibrium of the wire, and the proportion of conductive to radiative losses at these temperatures, it can be shown that our results are not affected by the maximum disturbance which could arise from these sources.

It remains to add a brief note on the connexion of this energy of "activation" with uses of the term in recent research and with implications outside our range of temperature The theory of Lennard-Jones (1932), explaining how energy might be needed for transition from molecular to atomic binding even when the work of dissociation is supplied by heat of adsorption, provided a picture of activation. Simultaneously the theory also suggested other activations needed for the migrations which allow diffusion of adsorbed gases into a metal. This second instance of activation was the one which obtained confirmation in experiments on slow sorption, to which accordingly the name of "activated adsorption" became commonly attached. Owing to the high velocity of "chemi-sorption", as the initial dissociation has often been called, the latter's status as activated adsorption has been lost in many publications, a situation which seems to call for direct measurements of temperature dependence of the condensation coefficient. For this our oscillographic method removes the obstacle of the high velocity.

In claiming that a law for oxygen deposition in this range of temperature and pressure has been thus brought within reach of experiment, we do not imply that it is the only law for all temperatures. For instance, if our estimates of K and E' are inserted in equation (1′), and this expression were assumed to be a complete account of the phenomena down to room temperature, there would be suggested a very slow rate of formation of the atomic layer from O_2 simply impinging upon cold tungsten without accumulating; on the other hand there exist several items of evidence suggesting that adsorptions indistinguishable from these in many of their consequences are able to take place without incandescence (e.g. Reimann 1935). It must, however, be remembered that the formation of an O_1 layer requires some source of O_2 and that at high and low temperatures the sources are not comparable, at our temperatures the large fraction of impinging molecules which fail to be activated will evaporate rapidly, so that no surface accumulation of O_2 can be considered as supplementing the supply impinging from the gas. But at low temperatures the alternatives available to molecules striking bare metal are not merely a dissociation or a desorption but also a finite duration of lifetime as surface layer within such lifetime will occur a slow transition into atomic states which is not allowed for in the simple treatment appropriate to our problem at high temperature.

5 SUMMARY

An experiment is devised for investigating the passage of small and reproducible quantities of oxygen over a hot tungsten filament whose surface purity can be ascertained by observation of its thermionic emission immediately before each admission of gas. The oxygen pressure at the filament surface is made to rise within a fraction of a second from a high vacuum to a transient value with a flat maximum of 10^{-3} to 10^{-4} mm. This enables the rate of primary adsorption to be isolated, and its dependence on gas pressure and on temperature to be determined, without the resulting monolayer becoming sufficiently dense to allow oxidation and appreciable thinning of the filament. The adsorption is traced by photographing the accompanying fall of electron emission by means of the oscillographic technique described in previous papers. Within a range of temperature and pressure such that evaporation is negligible and interaction between adsorbed particles is rare, the uncovered fraction of the surface decreases exponentially with time over a portion of each experiment. This enables a condensation constant to be obtained and its temperature dependence evaluated. The temperature coefficient which is found implies a heat of activation of 24,000 ± 3000 cal/mol. in the vicinity of 2200° abs. The bearing on theories of activated adsorption is briefly discussed.

REFERENCES

Johnson and Vick 1935 *Proc. Roy. Soc.* A, **151**, 296.
— — 1937*a* *Proc. Roy. Soc.* A, **158**, 55.
— — 1937*b* *Proc. Phys. Soc.* **49**, 409
Langmuir 1915 *J. Amer Chem. Soc.* **37**, 1161.
Lennard-Jones 1932 *Trans Faraday Soc.* **28**, 333.
Reimann 1935 *Phil. Mag* (7), **20**, 594.
Rideal and Wansbrough-Jones 1929 *Proc Roy. Soc.* A, **123**, 202.

Significance tests when several degrees of freedom arise simultaneously

By Harold Jeffreys, F.R.S.

(*Received* 29 *December* 1937)

1. If a set of observations are analysed for a new parameter a, which is initially as likely as not to be zero, and the possible range of whose values is s if it is not zero, we can denote the proposition that it is 0 by q, and the proposition that it is not 0 by $\sim q$ * Then the prior probabilities of q and $\sim q$ are given by

$$P(q \mid h) = P(\sim q \mid h) = \tfrac{1}{2}, \tag{1}$$

and the posterior probabilities on data θ are shown, by an approximate argument (Jeffreys 1937 *b*, p. 250), to be given by

$$K = \frac{P(q \mid \theta h)}{P(\sim q \mid \theta h)} \Big/ \frac{P(q \mid h)}{P(\sim q \mid h)} = \frac{s}{\sqrt{(2\pi)}\,\sigma_\alpha} \exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right), \tag{2}$$

where α is the maximum likelihood solution for a and σ_a its standard error. Since s is initially fixed and σ_α decreases like $n^{-\frac{1}{2}}$ when n, the number of observations, increases, the outside factor is proportional to $\sqrt{n}$. If K is less than 1, the observations support the introduction of the new parameter; if K is more than 1 they do not. In the cases so far examined the critical value of α/σ_α ranges from about 1·8 to 3 as the number of observations rises from 5 to 5000

1·1. Approximate rules can be given for cases where several parameters are tested at once. In such a case q is the proposition that none of them is needed, but $\sim q$ breaks up into a set of alternatives $q_1, q_2, \ldots, q_m$, each asserting the genuineness of one of the new parameters. These alternatives are not exclusive, since several of the new parameters may conceivably be relevant. In many cases, however, they can be supposed independent in the sense that the presence of one of them gives little or no reason to expect any of the others. Suppose then that the prior probability of any of $q_1, q_2, \ldots, q_m$ is k, and that these are independent Then the probability that all are false is $(1-k)^m$. But the proposition that all are false is q; hence if the presence of any of the new parameters is as likely as not we have

$$(1-k)^m = \tfrac{1}{2}, \tag{1}$$

whence

$$k = 1 - 2^{-1/m} \doteqdot m^{-1} \log 2, \tag{2}$$

---

* My q is always what Fisher (1935) calls a "null hypothesis".

when m is large. If we arrange our work so as to test one new parameter at a time we shall therefore have

$$\frac{P(\sim q_1 \mid h)}{P(q_1 \mid h)} = \frac{1-k}{k} = \frac{1}{2^{1/m}-1} \doteqdot \frac{m}{\log 2} = 1{\cdot}44\,m. \tag{3}$$

The ratio K depends wholly on the data, and the ratio of the posterior probabilities of $\sim q_1$ and q_1 will be got by multiplying the expressions 1 (2) and 1·1 (3). The test will therefore be the same as for one degree of freedom except that the outside factor must be multiplied by the quantity given in (3). It remains true that for constant m the outside factor is proportional to $n^{\frac{1}{2}}$.

1·2. It may happen that the propositions $q_1, q_2, \ldots, q_m$ are not independent; in some problems, if one new parameter is needed to express the data we should expect some of the others to be needed too. Thus, if a harmonic variation is suggested in a series of measures and a cosine term is present we may accept the corresponding sine term at once as giving only a determination of phase. In that case, if k is the probability that the cosine is needed, the probability that the sine is needed, given the cosine, is practically 1, and the probability that neither is needed is still $1-k$, not $(1-k)^2$. We should therefore get our test by taking

$$P(q \mid h) = P(q_1 \mid h) = \tfrac{1}{2},$$

and the test for a single degree of freedom needs no alteration.

Thus the rule for a single new parameter can be adapted easily to cases where several new parameters come into consideration together. The ratio $P(q \mid h)/P(q_1 \mid h)$ is the same whichever we choose to test first. In the latter type of case, we should naturally test the one that gives the larger coefficient in relation to its standard error, and the test would be the same as for one degree of freedom. If this term is a cosine one, and passes the test, the sine term will be accepted at once. In the former type, where the relevance of one new parameter gives no particular reason to expect any of the others, the allowance can be made by multiplying the outside factor by the expressions in 1·1 (3). The natural procedure would then be to compute χ^2 on the hypothesis that no new parameters are needed and inspect the distribution of the contributions to it. They can be tested in turn, beginning with the largest, and those that pass the test, with this allowance for selection, can be accepted as genuine.

1·3. There is an intermediate case, where the presence of one new parameter may increase the probability that another is needed without making its acceptance automatic. An approximation to this case can be made by an extension of Laplace's theory of sampling. Suppose that there are m elements, such as a system of means, with known standard errors, to be tested. Our problem is to say whether any of these show departures that support systematic differences, and if so, which. There is some difficulty in such a case about assessing the prior probability that the first one tested will be abnormal. I have on two previous occasions used $k = m^{-1}\log 2$, but this rests on the hypothesis of independence and is certainly too low. The difference is not very important, however, for practical application. Our problem will be to draw the line between the abnormal and normal cases, and the work will end at, say, the pth largest contribution to χ^2. Then at this stage $p-1$ abnormal contributions are known and $m-p$ normal ones The probability at this stage that the pth is abnormal is given by Laplace's theory as $p/(m+1)$, and we shall allow for it by multiplying the value of K by $(m-p+1)/p$. This will be very accurate. A convenient way of beginning is to accept as normal all contributions less than 1 and to take for the extreme one an additional factor got by putting $p = 1$, that is, simply m. This would be equivalent to saying that there is initially an even chance that one new parameter is needed, but that we do not know which of the m it is. If the extreme departure turns out to be significant with this allowance for selection, we can proceed inwards and outwards till the two series meet. At any stage the extra factor needed will be taken as $(r+1)/(s+1)$, where r and s are the numbers already accepted as normal and abnormal respectively It may happen that the number of small contributions to χ^2 is small; in that case a number of large ones will soon be accepted as systematic, and we may be led to proceed inwards so far as to cast doubt even on those originally taken as being random. Then these also will have to be reassessed Such a case may arise in some comparisons of measures of physical quantities where the systematic errors are suspected of being several times the admissible random ones.

1·4. We have therefore three different types of problem involving more than one degree of freedom, which can be treated by simple adaptations of the method for one degree. It appears to be important in any application to state clearly which type is the relevant one in the particular case studied. A test for an empirical period selected merely because it gives the largest amplitude of those found by harmonic analysis would be a case of 1·1 (3). In many agricultural experiments, I believe that the main tests refer to propositions whose prior probabilities could reasonably be taken as $\frac{1}{2}$, but the work is arranged to test various interactions of different degrees of complexity at the same time. The evidence for the relevance of the latter degrees of freedom would rest on the extreme departures from the predictions made by the hypothesis of independence, not on any previous ground for expecting them individually, even to the extent implied by a prior probability of $\frac{1}{2}$. In such a case the previous knowledge with respect to these higher degrees of freedom may be expressible by the rule 1·1 (3), which allows for selection.*

In the question of the mutual influence of earthquakes in different regions (1936, pp. 441–5) I used the rule of (1·1), but that of (1·3) would have been more appropriate. In discussing the systematic errors of seismological stations (1937 *a*, p. 39) I applied (1·1) to the extreme departure and then proceeded according to (1·3), but it would have been better to use 1·3 throughout. The difference does not affect the inferences actually drawn.

2 Now even with such allowance for selection as may be needed it remains true that the outside factor in the support for q is of order $n^{\frac{1}{2}}$, this factor would be the support provided if the estimates happened to agree exactly with the predictions made by q. In the case where the occurrence of one new function in a representation would imply one or more others, as for a harmonic variation, the test for all together will approximate to that for the largest coefficient. But this condition is not satisfied by the test that I have previously given for such a case (1936): if the number of functions entering together is m, I obtained a factor $n^{\frac{1}{2}m}$. It seems that this cannot be right. The method used treated them symmetrically, which was desirable, but if a cosine term has coefficient 3 ± 1 it can hardly make much difference to its acceptance whether the sine has coefficient 0 ± 1 or 1 ± 1, and this only makes a difference of a factor 0·6 to K in my previous formula. Within such limits the test for a cosine separately should be correct for a cosine and sine together, and if the coefficient is significant both terms should be accepted. The essential point is that the outside factor, for any value of m, must be of order $n^{\frac{1}{2}}$. If it was of order $n^{\frac{1}{2}m}$ this approximation by testing the largest coefficient would be utterly wrong, and the considerations given here show that apart from a factor of order unity it should be right for all values of n. Indeed the factor n for a pair of harmonic terms makes the test for the two together as severe as separate tests for the two would be, disregarding the fundamental condition of the problem that if one is accepted the other will be.

The factor $n^{\frac{1}{2}}$ in the problem of one new parameter enters because if the solution for a is under σ_a, that is what might be expected if a was really 0, but if a might have been anywhere in a range s the probability of the result would have been only of order σ_a/s. Hence we accept the hypothesis that requires the less remarkable coincidence, and our rules show that in these conditions $P(q \mid \theta h)/P(\sim q \mid \theta h)$ must be of order s/σ_a, which is of order $n^{\frac{1}{2}}$. The appearance of $n^{\frac{1}{2}m}$ in the test of several new parameters means that if they all come out less than their standard errors it requires m coincidences of this kind on the hypothesis that they are genuine. It is clear, however, that it only needs one. If the amplitude of a harmonic term is small both coefficients must be. Further, if the prior probability of an amplitude, on hypothesis $\sim q$, is taken as uniformly distributed from 0 to s, and the amplitude found is of order σ_a, the ratio K will be of order $n^{\frac{1}{2}}$, which is what we need for consistency with the approximate rule. Any other power of the amplitude would lead to a different power of n.

---

* Fisher (1936, pp. 65–6) makes an analogous recommendation.

The rule given for the prior probability for the case of a pair of harmonic terms $\sqrt{2}(a\cos x + b\sin x)$ was, for $a^2+b^2<s^2$,

$$P(da\,db \mid \sim q, s, h) = da\,db/\pi s^2, \tag{1}$$

where s^2 is the true expectation of the outstanding variation when we begin to consider the terms. This says that if a and b are small compared with s they give a negligible amount of information about each other. But this is plainly wrong. In the conditions considered, which will regard the phase as initially unknown, we should normally expect b to be of the same order of magnitude as a until there was positive reason to believe the contrary. If a was small and the phase happened by a coincidence to be nearly 0 or $\frac{1}{2}\pi$ this would not be true, but this has a negligible prior probability if the phase is initially unknown. The above form would say, in fact, that if the coefficient of the cosine is small the probability of that of the sine is still uniformly distributed from $-\sqrt{(s^2-a^2)}$ to $+\sqrt{(s^2-a^2)}$, and therefore that we can be nearly sure that the phase is nearly $\pm\frac{1}{2}\pi$. This is not the inference that anybody would draw, he would infer that the amplitude was probably small and that the phase might still be almost anywhere in a range π.

The form (1) was adopted because it would make the posterior probability density for a and b, given s, come at the maximum likelihood solution, and therefore was in accordance with present practice. This consideration, however, is irrelevant, because with any ordinary distribution of the prior probability, uniform or not, the estimates by inverse probability and maximum likelihood would agree within much less than the standard error of a determination by either method (Jeffreys 1938*a*); the only exception would be if our previous information about a and b was comparable in completeness with what the new data could tell us This does not arise in the present problem, in which their very existence is in doubt. It is clear that the hypothesis of the prior irrelevance of a and b to each other must be abandoned. The objection to it is analogous to the correction that I have made already to my previous solutions for sampling and contingency (1937 *d*). In this I showed that for two adjacent entries in a 2×2 classification to be both small we require only one coincidence, whereas my previous solution, assuming too great a degree of independence, would require two. The same applies to a pair of Fourier coefficients.

To bring the analysis into accordance with the above considerations we must take the prior probability of the amplitude $c = \sqrt{(a^2+b^2)}$ as uniformly distributed from 0 to s. (If we combined the uniform distribution of the prior probability of c^2 with uniform distribution for phase we should be led back to (1) and to the previous inconsistencies.) Then

$$P(dc \mid \sim q, s, h) = dc/s. \tag{2}$$

The phase ϕ may be from 0 to 2π, hence if the amplitude gives no information about the phase

$$P(d\phi \mid \sim q, s, c, h) = d\phi/2\pi \tag{3}$$

and $$P(dc, d\phi \mid \sim q, s, h) = \frac{dc\,d\phi}{2\pi s} = \frac{da\,db}{2\pi cs} = P(da\,db \mid \sim q, s, h). \tag{4}$$

This form will replace (1). The joint probability density of a and b is no longer uniform, since it increases when c is small.

To generalize to m degrees of freedom, we seek a representation in the form

$$y = Sa_l f_l(x), \tag{5}$$

where S denotes summation over all the functions f_l and the functions are orthogonal. We have no previous opinion on the relative importance of the m functions f_l, and shall therefore maintain symmetry between them. We take them so that the expectation of f_l^2 when x is equally likely to be anywhere in its permitted range is 1, thus for a constant term we should take $f_0 = 1$, for a linear term, valid between $x = \pm 1$, we should take $f_1 = x\sqrt{3}$, for a pair of harmonic terms we should take $f_1 = \sqrt{2}\cos x$, $f_2 = \sqrt{2}\sin x$. Then the expectation of the contribution of the new terms to y^2 is Sa_l^2, by the condition of orthogonality. We put

$$Sa_l^2 = c^2; \tag{6}$$

c is then a generalized amplitude for the new terms. If it is small it implies at once that all the new terms are small. The hypothesis $\sim q$ asserts that c is not zero, and the case of two degrees of freedom suggests that we should take

$$P(dc \mid \sim q, s, h) = dc/s, \tag{7}$$

if s is the greatest value that c can have. The coefficients a_l may be regarded as the components of a vector of length c in m dimensions, whose direction is unknown. Given that this vector is between c and $c+dc$, we express ignorance of its direction by taking

$$P(da_1 da_2 \dots da_m \mid \sim q, dc, s, h) = da_1 da_2 \dots da_m / dV, \tag{8}$$

where dV is the volume of the spherical shell between radii c and $c+dc$. The volume of an m-dimensional sphere of radius c is

$$V = \frac{\pi^{\frac{1}{2}m} c^m}{(\frac{1}{2}m)!}, \tag{9}$$

whence
$$dV = \frac{2\pi^{\frac{1}{2}m} c^{m-1}}{(\frac{1}{2}m-1)!} dc. \tag{10}$$

Hence, by the product formula,

$$P(da_1 da_2 \dots da_m \mid \sim q, s, h) = \frac{(\frac{1}{2}m-1)!}{2\pi^{\frac{1}{2}m} c^{m-1} s} da_1 da_2 \dots da_m. \tag{11}$$

This is easily seen to reduce to the correct forms $da/2s$ and $da\,db/2\pi cs$ when $m=1$ and $m=2$.

3. *Lemma* We need an approximation to the integral

$$I = \iint \dots \int (Sa_l^2)^{-\frac{1}{2}(m-1)} \exp\{-k^2 S(a_l - \alpha_l)^2\} \, da_1 \dots da_m \tag{1}$$

over all values of the a's, S denoting summation with regard to l from 1 to m. First change the variables to an orthogonal set b, so that

$$b_1 = \frac{S\alpha_l a_l}{(S\alpha_l^2)^{\frac{1}{2}}}. \tag{2}$$

Then b_1 is the component of the vector a_l in m dimensions in the direction of α_l, while the other b's are components at right angles to b_1 and to one another. Then the integral is

$$I = \iint \dots \int (Sb_l^2)^{-\frac{1}{2}(m-1)} \exp\{-k^2 (b_1 - \sqrt{S\alpha_l^2})^2 - k^2 S' b^2\} \, db_1 \dots db_m, \tag{3}$$

S' denoting that b_1 is omitted from the summation. Now put

$$b_2^2 + \dots + b_m^2 = f^2. \tag{4}$$

These variables appear only through f; and through a volume between two slightly different values of f,

$$\int \dots \int db_2 \dots db_m = \frac{2\pi^{\frac{1}{2}(m-1)} f^{m-2}}{(\frac{1}{2}m-\frac{1}{2})!} df. \tag{5}$$

Thus

$$I = \iint (b_1^2+f^2)^{-\frac{1}{2}(m-1)} \exp\{-k^2(b_1-\sqrt{S\alpha^2})^2-k^2f^2\} \frac{2\pi^{\frac{1}{2}(m-1)}}{(\frac{1}{2}m-\frac{3}{2})!} f^{m-2}\,db_1\,df. \tag{6}$$

Now transform the variables so that

$$b_1^2+f^2 = c^2, \quad f^2 = c^2t. \tag{7}$$

Then

$$\frac{\partial(b_1,f)}{\partial(c,t)} = \frac{c}{2t^{\frac{1}{2}}(1-t)^{\frac{1}{2}}}, \tag{8}$$

$$I = \frac{\pi^{\frac{1}{2}m}}{(\frac{1}{2}m-\frac{3}{2})!\,k}\int_0^1 t^{\frac{1}{2}m-\frac{3}{2}}(1-t)^{-\frac{1}{2}}\exp(-k^2t\,S\alpha^2)\,dt \tag{9}$$

So far this is exact.* It is necessary that $m \geqslant 2$ for the previous transformations to be possible. The integral is therefore always convergent. For $m = 2$ the integrand is infinite at $t = 0$, for $m = 3$ it is finite, and for larger values of m it is zero there.

Consider first the case $S\alpha^2 = 0$. The integral is then a Beta function, and

$$I = \frac{\pi^{\frac{1}{2}m+\frac{1}{2}}}{(\frac{1}{2}m-1)!\,k}. \tag{10}$$

I decreases steadily as $S\alpha^2$ increases. For brevity we denote the latter simply by S. Suppose now that k^2S is large, so that most of the integral comes from fairly small values of t. If we omit the factor in $1-t$, the maximum of the integrand is at

$$t = \frac{m-3}{2k^2S} \tag{11}$$

for $m > 3$; for $m = 2$ or 3 the extreme value is at $t = 0$. For $2k^2S = m$ the maximum is at $t < \frac{1}{2}$ until $m = 6$. The important values in our applications will be for $2k^2S$ somewhat greater than m, and the largest value of m that appears worth considering is 10, so that the maximum will never be near 1. Now for $t = \frac{1}{2}$

$$(1-t)^{\frac{1}{2}}\exp(\tfrac{1}{2}t) = 2^{-\frac{1}{2}}e^{\frac{1}{4}} = 0{\cdot}908 \tag{12}$$

and the difference from 1 for smaller values of t is nearly proportional to t^2. So long as the maximum is at a value of t less than $\frac{1}{4}$ we can therefore replace $(1-t)^{-\frac{1}{2}}$ by $\exp(\frac{1}{2}t)$. For larger values of m, where the maximum may be

---

* The transformation has a branch point at $b_1 = 0$ $(t = 1)$ and the integral is the sum of two corresponding to opposite signs of b_1. If we restrict ourselves to cases where the observed amplitude has the right sign, another factor

$$\tfrac{1}{2}\{1+\operatorname{erf} k\sqrt{(S\alpha^2)}\sqrt{(1-t)}\}$$

is needed, but this is unimportant in actual cases.

near $\frac{1}{2}$, the maximum is sharp and we may replace $1-t$ by its value at the maximum. With either approximation we can extend the range to infinity.

With the former approximation,

$$I = \frac{\pi^{\frac{1}{2}m}}{(\frac{1}{2}m-\frac{3}{2})!\,k}\int_0^\infty t^{\frac{1}{2}m-\frac{3}{2}}\exp\{(-k^2S+\tfrac{1}{2})t\}\,dt$$

$$= \frac{\pi^{\frac{1}{2}m}}{k(k^2S-\frac{1}{2})^{\frac{1}{2}m-\frac{1}{2}}}. \tag{13}$$

This will be valid for $m = 2$ and 3, and for larger values if $2k^2S$ is larger than $4(m-3)$.

With the latter,

$$I = \frac{\pi^{\frac{1}{2}m}}{(\frac{1}{2}m-\frac{3}{2})!\,k}\int_0^\infty t^{\frac{1}{2}m-\frac{3}{2}}\exp(-k^2St)\,dt\,.\left(1-\frac{m-3}{2k^2S}\right)^{-\frac{1}{2}}$$

$$= \frac{\pi^{\frac{1}{2}m}}{k(k^2Sa^2)^{\frac{1}{2}m-\frac{1}{2}}}\left(1-\frac{m-3}{2k^2S}\right)^{-\frac{1}{2}}. \tag{14}$$

This will be valid for $m = 4$ and more if $2k^2S$ is larger than m and between $2(m-3)$ and $4(m-3)$

A different form of approximation would be needed if the maximum came at $t = \frac{2}{3}$ or more, but this does not concern us If $m = 10$ and $2k^2S = m$, t is still only 0·7, and it will be seen that the relevant values of $2k^2S$ are distinctly larger.

If $m = 1$, (1) integrates directly and gives

$$I = \pi^{\frac{1}{2}}/k \tag{15}$$

for all values of S. But (10) and (13) reduce to this for $m = 1$ and are therefore valid in this case also.

If we form I_0 and I_m, the values taken by I in (10) and (13) when $2k^2S$ is 0 and m, we have

$$\frac{I_m}{I_0} = \frac{(\frac{1}{2}m-1)!}{\pi^{\frac{1}{2}}(\frac{1}{2}m-\frac{1}{2})^{\frac{1}{2}m-\frac{1}{2}}} \doteq (2e)^{\frac{1}{2}}e^{-\frac{1}{2}m}. \tag{16}$$

The approximation from (14) is more doubtful but suggests a ratio of order $(\frac{2}{3}m)^{\frac{1}{2}}e^{-\frac{1}{2}m}$.

We could have proceeded by integrating (1) directly with regard to all the variables, giving to Sa^2 its value at the maximum of the exponential, namely, $S\alpha^2$ This gives immediately

$$I = \frac{\pi^{\frac{1}{2}m}}{k^m(S\alpha^2)^{\frac{1}{2}(m-1)}}. \tag{17}$$

(13) differs by the subtraction of $\frac{1}{2}$ from $k^2 S\alpha^2$, and (17) is therefore a little too small. This method of course breaks down completely for small values of $S\alpha^2$, since it is then impossible to neglect the variation of $S\alpha^2$ on the various paths of integration. This makes the integrand vary more rapidly than the exponential for all directions except b_1. For larger values of $S\alpha^2$ the effect is still not quite negligible, but the effect of the variation with b_1 is slightly more important than the aggregate of the others It was on account of the uncertainty of the correction for this effect that the present method of exact integration for $m-1$ of the components was devised.

4. *Departure from a uniform distribution of chance.* This is the easiest case to discuss further. As for the case already treated (1938*b*), where the suggested departure was linear, we introduce a parameter t linearly related to the independent variable x, and equal to 0 and 1 at the upper and lower limits of x Then on the hypothesis of a uniform distribution, which we denote by q, the chance of an observation lying in a particular range is dt. On the hypothesis of a departure of the type suggested the corresponding chance is $\{1+Sa_l f_l(t)\}\,dt$ (S denoting summation over the different functions $f_l(t)$). We take the two hypotheses as initially equally probable; and if

$$Sa_l^2 = c^2, \tag{1}$$

and the maximum possible value of c is s, we can write

$$P(q\,|\,h) = \tfrac{1}{2},\ \ P(\theta\,|\,qh) = \Pi(dt), \tag{2}$$

$$P(\sim q\,|\,h) = \tfrac{1}{2},\ \ P(da_1 \ldots da_m\,|\sim q, h) = \frac{(\frac{1}{2}m-1)!}{2\pi^{\frac{1}{2}m}c^{m-1}s}\,da_1\,da_2 \ldots da_m, \tag{3}$$

$$P(\sim q\,da_1 \ldots da_m\,|\,h) = P(\sim q\,|\,h)\,P(da_1\,..\,da_m\,|\sim q, h), \tag{4}$$

$$P(\theta\,|\sim q\,da_1\ .\,da_m, h) = \Pi\{1+Sa_l f_l(t)\}\,dt, \tag{5}$$

whence

$$P(q\,|\,\theta h) \propto 1, \tag{6}$$

$$P(\sim q\,|\,\theta h) \propto \frac{(\frac{1}{2}m-1)!}{2\pi^{\frac{1}{2}m}s}\iint\ldots\int\frac{1}{c^{m-1}}\Pi\{1+Sa_l f_l(t)\}\,da_1\,da_2\ ..\,da_m. \tag{7}$$

Now if we write
$$\phi = \Sigma\log\{1+Sa_l f_l(t)\} \tag{8}$$

(Σ denoting summation over the observations), we have

$$\frac{\partial\phi}{\partial a_k} = \Sigma\frac{f_k(t)}{1+Sa_l f_l(t)} \doteq \Sigma f_k(t) - \Sigma f_k(t)\,Sa_l f_l(t), \tag{9}$$

$$\frac{\partial^2\phi}{\partial a_k^2} = -\Sigma\frac{f_k^2(t)}{\{1+Sa_l f_l(t)\}^2};\ \ \frac{\partial^2\phi}{\partial a_k \partial a_p} = -\Sigma\frac{f_k(t)\,f_p(t)}{\{1+Sa_l f_l(t)\}^2}. \tag{10}$$

Now it has been supposed that the functions are orthogonal and that the integral of the square of each is 1. It is also supposed that the coefficients a_l are small in the actual case. Then the sum of $f_l^2(t)$ over the observations, whose number is n, will be nearly n, and that of $f_l(t)f_k(t)$ will be much smaller on account of the condition of orthogonality. Then the stationary value of ϕ will be nearly at

$$a_l = \alpha_l = \frac{1}{n}\Sigma f_l(t), \tag{11}$$

and

$$\phi = \Sigma\log\{1+S\alpha_l f_l(t)\} - \tfrac{1}{2}nS(a_l-\alpha_l)^2. \tag{12}$$

But $\phi = 0$ when all the $a_l = 0$, and can therefore be written

$$\phi = \tfrac{1}{2}nS\alpha_l^2 - \tfrac{1}{2}nS(a_l-\alpha_l)^2 \tag{13}$$

to order a^2. Then

$$P(\sim q\,|\,\theta h) \propto \frac{(\frac{1}{2}m-1)!}{2\pi^{\frac{1}{2}m}s}\iint\ldots\int (Sa_l^2)^{\frac{1}{2}(m-1)}$$
$$\times\exp(\tfrac{1}{2}nS\alpha_l^2)\exp\{-\tfrac{1}{2}nS(a_l-\alpha_l)^2\}\,da_1\ldots da_m. \tag{14}$$

The integral has the form treated in the lemma, with $k^2 = \frac{1}{2}n$, apart from the fact that the range is of order 1 instead of infinity; but both factors decrease so rapidly that this difference can be neglected. Hence for $S\alpha^2 = 0$ we have

$$P(\sim q\,|\,\theta h) \propto (\pi/2n)^{\frac{1}{2}} \tag{15}$$

for all values of m.

The standard error of any α being $n^{-\frac{1}{2}}$, we have

$$nS\alpha^2 = \chi^2,\quad k^2S\alpha^2 = \tfrac{1}{2}\chi^2, \tag{16}$$

and for $m = 2$ or 3 and $\chi^2 > m$, or $m > 3$, $\chi^2 > 4(m-3)$,

$$P(\sim q\,|\,\theta h) \propto \frac{2^{\frac{1}{2}m-1}(\frac{1}{2}m-1)!}{n^{\frac{1}{2}}s}\frac{e^{\frac{1}{2}\chi^2}}{(\chi^2-1)^{\frac{1}{2}(m-1)}}, \tag{17}$$

which is also right for $m = 1$.

For $m > 3$, $m < \chi^2$, $2(m-3) < \chi^2 < 4(m-3)$,

$$P(\sim q\,|\,\theta h) \propto \frac{2^{\frac{1}{2}m-1}(\frac{1}{2}m-1)!}{n^{\frac{1}{2}}s}\frac{e^{\frac{1}{2}\chi^2}}{\chi^{m-1}}\left(1-\frac{m-3}{\chi^2}\right)^{-\frac{1}{2}}. \tag{18}$$

Since $P(q\,|\,\theta h) \propto 1$, K is in each case the reciprocal of the expression (15), (17), or (18)

K will evidently contain the factor $\chi^{m-1}\exp(-\frac{1}{2}\chi^2)$, which is exactly the function that occurs in the usual χ^2 test. The other factors are either constant or vary much more slowly with χ^2. They really arise because χ^2 is not linearly related to any of the unknowns. For one degree of freedom, the estimate is unbiased and the extra factor does not appear. But for any larger number of degrees of freedom $\chi/\sqrt{n}$ is an estimate of the resultant amplitude systematically increased by the sampling error. What the test does, effectively, is to allow for this increase by reducing χ^2 slightly.

Another way of expressing the point is to assume from the start that the function of the observations relevant to the test is χ^2. Then on hypothesis q, χ^2 is entirely due to random error and the probability that it would be in a given range is proportional to $\chi^{m-1}\exp(-\frac{1}{2}\chi^2)\,d\chi$. On $\sim q$, if the real amplitude is large compared with the random errors, χ is nearly proportional to the amplitude, and as we have taken the prior probability of the latter uniform the probability of $d\chi$ is proportional to $d\chi$. Comparison gives K proportional to $\chi^{m-1}\exp(-\frac{1}{2}\chi^2)$. But on $\sim q$, if the real amplitude is small, the random variation still adds something near m to χ^2, and this produces an outward displacement of the estimated coefficients and an additional concentration of χ^2 in the region just above m. Hence in this region the probability of $d\chi$, given $\sim q$, is distributed a little more densely than for larger values, and the function of χ^2 that appears in K will be a little less than $\chi^{m-1}\exp(-\frac{1}{2}\chi^2)$.

When $\chi^2 = m$, and $m = 2$ or 3, K is about $(2e)^{-\frac{1}{2}}$ of its value for $\chi^2 = 0$, and therefore is still large if n is large. For larger values of m the value of K at $\chi^2 = m$ is smaller, at least if the approximation given is valid; but the critical values are substantially larger than m and I have not thought it worth while to examine the appropriate approximation more closely.

It is satisfactory that n enters through the power $n^{\frac{1}{2}}$ If we had chosen a different power of c in (3) a different power of n would have appeared in the result. The elementary argument based on testing the extreme coefficient by itself must give the right order of magnitude for the result, and leads to this power of n. The advantage of a more detailed discussion such as this is that it uses the whole of the relevant information contained in the observations, whereas the simpler one does not. Thus if the coefficient of a cosine term came out $0{\cdot}3 \pm 0{\cdot}1$, this discussion would make a difference according as that of the sine was $0{\cdot}1$ or $0{\cdot}25$, the elementary one would not. But it is reasonable to expect that the elementary theory would give a result of the right order of magnitude for all numbers of observations, and this condition is satisfied.

The test will therefore say at what value of χ^2, greater than m, the data begin to support the introduction of new functions into the chance.

Apart from the cases $m = 1$ and 2 I know of no practical applications where the conditions required for this test are satisfied. The case $m = 1$ has already been treated. A harmonic variation is an instance where $m = 2$. If

$$f_1(t) = \sqrt{2}\cos 2\pi t; \quad f_2(t) = \sqrt{2}\sin 2\pi t,$$

the extreme values of $a_1 f_1 + a_2 f_2$ are $\pm\sqrt{2}(a_1^2+a_2^2)^{\frac{1}{2}}$. Since $1+Sa_l f_l$ must be everywhere positive the extreme value of c is $1/\sqrt{2}$, which is therefore the appropriate value of s.

It is possible, however, that the true extreme would be a case where the entire chance is concentrated in one half-period, within which we can treat it as uniformly distributed. This would require higher harmonics to express it, but we may reasonably suppose that their coefficients are proportional to higher powers of a_1 and a_2, and would not affect the prior probability distribution of the latter. Then if the true probability of an observation in a range dt is $2dt$ from $t = 0$ to $\frac{1}{2}$, 0 from $\frac{1}{2}$ to 1, we find by Fourier analysis

$$a_1 = 0, \ a_2 = 2\sqrt{2}/\pi,$$

whence in a case of this type we should take $s = 2\sqrt{2}/\pi$

If the frequency distribution of stars with regard to direction from the sun was less obviously non-uniform, this test could be used to test the presence of a departure expressible by the spherical harmonics of a given order, so that higher values of m would arise There is nothing in the argument that depends on the restriction of $f_l(t)$ to be functions of only one variable.

5. *The representation of measurements by assigned functions* The fundamental difference between this problem and that of testing a law of chance is that the random variation might turn out indefinitely small, in the problem of chances it is fixed by the numbers in the samples themselves. Thus three different observations lying exactly on a straight line would be strong evidence against constancy; but three like events in succession would not be strong evidence against an even chance. The systematic variation might account for any fraction of the whole variation outstanding at the beginning of the investigation, and the generalized amplitude c may reach s, the square root of the expectation of the square of a residual outstanding at that stage. But s is to be found from the scatter of the observations and is not, as in problems of sampling or distributions of chance, determinable in advance. It therefore enters as an unknown with the prior probability distribution appropriate to an unknown essentially positive quantity. This consideration was taken into account previously, but the distribution of the prior probabilities of the coefficients of the new functions that might arise, for given s, was wrong for the reasons indicated above.

An unconsidered complication also needs attention. In my previous discussion (1936, pp 432–7) I took the new functions f_l to be mutually orthogonal at the actual weights. This leads to a mathematical simplification, but otherwise has little in its favour. It may well happen, in the case of a suggested harmonic variation, that the bulk of the observations are concentrated near $x = 0$ and $x = \pi$, so that they give a good determination of the coefficient of a cosine but a very bad one of the sine. If we are in a position to choose the values of x where we make our observations of the quantity y that we are trying to represent, we should naturally place them to give equally good determinations of both, and the condition of orthogonality at the actual weights would be satisfied. Anybody concerned mainly with natural phenomena rather than experimental ones will, however, have to deal with the observations as they occur, and this condition may be far from being satisfied. It is therefore desirable to extend the analysis to allow for this complication. The statement of the problem is then as follows:

$$P(q, ds \mid h) \propto ds/s, \tag{1}$$

$$P(\sim q, ds, da_1 \, . \; da_m \mid h) \propto \frac{(\frac{1}{2}m-1)!}{2\pi^{\frac{1}{2}m}(Sa_l^2)^{\frac{1}{2}(m-1)}} \frac{ds}{s^2} da_1 \, . \; da_m, \tag{2}$$

and if the observed values of y are denoted collectively by θ,

$$P(\theta \mid q, ds, h) = \frac{1}{(2\pi s^2)^{\frac{1}{2}n}} \exp\left(-\Sigma \frac{y^2}{2s^2}\right) \Pi(dy), \tag{3}$$

$$P(\theta \mid \sim q, ds, da_1 \ldots da_m, h) = \frac{1}{\{2\pi(s^2 - Sa^2)\}^{\frac{1}{2}n}} \exp\left(-\Sigma \frac{(y - Sa_l f_l)^2}{2(s^2 - Sa^2)}\right) \Pi(dy) \tag{4}$$

Hence

$$P(q, ds \mid \theta h) \propto \frac{ds}{s^{n+1}} \exp\left(-\Sigma \frac{y^2}{2s^2}\right), \tag{5}$$

$$P(\sim q, ds, da_1 \ldots da_m \mid \theta h)$$
$$\propto \frac{(\frac{1}{2}m-1)! \, ds \, da_1 \ldots da_m}{2\pi^{\frac{1}{2}m}(Sa^2)^{\frac{1}{2}(m-1)}(s^2 - Sa^2)^{\frac{1}{2}n} s^2} \exp\left(-\Sigma \frac{(y - Sa_l f_l)^2}{2(s^2 - Sa^2)}\right). \tag{6}$$

Put in (5) $$\Sigma y^2 = n\sigma^2; \; s^2 = 1/2k^2. \tag{7}$$

The exponential in (6) is the usual likelihood factor. Its maximum is at $a_l = \alpha_l$, the solution of the normal equations, and we may write in (6)

$$a_l = \alpha_l + \xi_l; \; s^2 - Sa^2 = 1/2k^2; \; \Sigma(y - Sa_l f_l)^2 = n\sigma'^2. \tag{8}$$

Then $$\Sigma(y - Sa_l f_l)^2 = \Sigma(y - S\alpha_l f_l)^2 + \Sigma(S\xi_l f_l)^2$$
$$= n\sigma'^2 + \Sigma(Sf_l \xi_l)^2. \tag{9}$$

If the observations are so distributed that the products of the ξ_l in (9) cancel we have the condition of orthogonality. But consider first the opposite case, where the discriminant of this expression is so small that most of the variation of the integrand with regard to a_m comes from that of Sa^2 and not from the terms in a_m in the exponential. In this case the maximum with regard to a_m is at $a_m = 0$, and if we put

$$Sa^2 = S'a^2 + a_m^2, \tag{10}$$

we have
$$\int (Sa^2)^{-\frac{1}{2}(m-1)}\, da_m = \frac{\pi^{\frac{1}{2}}(\frac{1}{2}m-2)!}{(\frac{1}{2}m-\frac{3}{2})!}(S'a^2)^{-\frac{1}{2}(m-2)}, \tag{11}$$

and we can put $a_m = 0$ in the other factors in (6). Then (6) becomes

$$P(\sim q, ds, da_1 \ldots da_{m-1} \mid \theta h)$$
$$\propto \frac{(\frac{1}{2}m-1)!\,(\frac{1}{2}m-2)!}{2\pi^{\frac{1}{2}(m-1)}(\frac{1}{2}m-\frac{3}{2})!}\, \frac{ds\, da_1 \,..\, da_{m-1}}{(S'a^2)^{\frac{1}{2}(m-2)}(s^2-S'a^2)^{\frac{1}{2}n}s^2} \exp\left[-\Sigma\frac{(y-S'a_l f_l)^2}{2(s^2-S'a^2)}\right]. \tag{12}$$

But if we ignored a_m from the start, and considered only the $m-1$ parameters a_1 to a_{m-1}, (6) would be replaced by

$$P(\sim q, ds, da_1 \ldots da_{m-1} \mid \theta h)$$
$$\propto \frac{(\frac{1}{2}m-\frac{3}{2})!\, ds\, da_1 \ldots da_{m-1}}{2\pi^{\frac{1}{2}(m-1)}(S'a^2)^{\frac{1}{2}(m-2)}(s^2-S'a^2)^{\frac{1}{2}n}s^2} \exp\left[-\Sigma\frac{(y-S'a_l f_l)^2}{2(s^2-S'a^2)}\right], \tag{13}$$

and the ratio of these two expressions is

$$\frac{\{(\frac{1}{2}m-\frac{3}{2})!\}^2}{(\frac{1}{2}m-1)!\,(\frac{1}{2}m-2)!} = 1 + O\left(\frac{1}{m}\right). \tag{14}$$

Thus the two expressions are equivalent for large m. For moderate values we have.

m	2	3	4	5	6	7	8	9	10
Ratio	0	0·64	0·78	0·85	0·89	0·91	0·92	0·93	0·94

the expression in (13) being the smaller. In such cases we should therefore proceed by rejecting the worst-determined parameter altogether, test the remainder by (13), and allow for the rejection by dividing $P(\sim q \mid \theta h)$ by the appropriate number in the table; and therefore K will be multiplied by the same number. For $m = 2$ the exponential factor is still necessary to save convergence and the approximation (11) fails. In this case we cannot reject an unknown.

The method suggested, if great departure from orthogonality is detected in the solution of the normal equations, will then be as follows. Transform the a_l to a new set of orthogonal parameters b_l (not the b_l of the Lemma) such that the product terms of the exponent vanish. Then the variation with regard to b_m due to Sb^2 will be represented by a factor $\frac{1}{2}(m-1)\,b_m^2/S'b^2$, and that due to the exponential will be represented by $-\frac{1}{2}b_m^2/\sigma^2(b_m)$. The former will be the more important if

$$\sigma^2(b_m) > S'b^2/(m-1), \tag{15}$$

and therefore if the standard error of b_m exceeds the root mean square of the estimates of the other parameters. If this is found to be true it will be best to ignore b_m and test the other parameters similarly, applying the correcting factor (14) in each case. The reduction of K will be partly cancelled by the disappearance of one term from χ^2

The tendency of this rejection of badly determined parameters is to improve the orthogonality of the set, especially in the cases where the estimates are such that there may be any doubt about significance. Consequently we may proceed to consider the problem as if orthogonality held. Then (6) reduces to

$$\begin{aligned} P(\sim q, ds, da_1 \ldots da_m \mid \theta h) &\propto \frac{(\frac{1}{2}m-1)!\,ds\,da_1\ldots da_m}{2\pi^{\frac{1}{2}m}(S\alpha^2)^{\frac{1}{2}(m-1)}(s^2-S\alpha^2)^{\frac{1}{2}n}\,s^2} \\ &\qquad\times\exp[-nk^2\{\sigma'^2+S(a_l-\alpha_l)^2\}] \\ &\propto \frac{(\frac{1}{2}m-1)!\,ds\,da_1\ldots da_m}{2\pi^{\frac{1}{2}m}(S\alpha^2)^{\frac{1}{2}(m-1)}\,s^2}\,k^n \\ &\qquad\times\exp[-nk^2\{\sigma'^2+S(a_l-\alpha_l)^2\}], \end{aligned} \tag{16}$$

(5) at the same time, the same factor being dropped, becoming

$$P(q, ds \mid \theta h) \propto \frac{ds}{s}\,k^n \exp-(nk^2\sigma^2). \tag{17}$$

The integration with regard to the a_l can then be done by the lemma. In the first place, if all the α_l are 0, we find

$$P(\sim q, ds \mid \theta h) \propto \left(\frac{\pi}{2n}\right)^{\frac{1}{2}} \frac{k^n}{s}\,e^{-nk^2\sigma'^2}ds, \tag{18}$$

and in this case $\sigma' = \sigma$. Comparing with (17) we have

$$K = (2n/\pi)^{\frac{1}{2}}, \tag{19}$$

as for one degree of freedom and as for the frequency problem.

For $n = 2$ or 3, and $2nk^2S\alpha^2 > m$, or $n \geqslant 4$ and $2nk^2S\alpha^2 > 4(m-3)$,

$$P(\sim q, ds \mid \theta h) \propto \frac{(\frac{1}{2}m-1)!}{2k\sqrt{n}}\,\frac{k^n ds}{s^2}\,\frac{\exp(-nk^2\sigma'^2)}{(nk^2S\alpha^2-\frac{1}{2})^{\frac{1}{2}(m-1)}}. \tag{20}$$

For $n > 3$ and $2nk^2S\alpha^2 > m$, $2(m-3) < 2nk^2S\alpha^2 < 4(m-3)$,

$$P(\sim q, ds\,|\,\theta h) \propto \frac{(\frac{1}{2}m-1)!}{2k\sqrt{n}}\,\frac{k^n ds}{s^2}\,\frac{\exp(-nk^2\sigma'^2)}{(nk^2S\alpha^2)^{\frac{1}{2}(m-1)}}\left(1-\frac{m-3}{2nk^2S\alpha^2}\right)^{-\frac{1}{2}}. \tag{21}$$

It remains to integrate with regard to s The rapidly varying factors in (17) are $k^n \exp(-nk^2\sigma^2)$ and in (20) and (21) are $k^{n-m}\exp(-nk^2\sigma'^2)$. It turns out that it is not necessarily legitimate to neglect the variation of k^{-m} in comparison with the other factors. The maxima are at $k^2 = 1/2\sigma^2$ and $k^2 = (n-m)/2n\sigma'^2$. Hence our estimate of s in (17) is σ, but in (20) and (21)

$$\sigma'^2 = \sigma^2 - S\alpha^2, \tag{22}$$

$$s^2 - S\alpha^2 = \frac{n\sigma'^2}{n-m}. \tag{23}$$

The usual estimate of the standard error of α_l is $\sigma'/\sqrt{(n-m)}$. Then

$$\chi^2 = \frac{(n-m)\,S\alpha^2}{\sigma'^2}. \tag{24}$$

Integrating, we have, from the main factors in (20) and (21),

$$K = \frac{n^{\frac{1}{2}n}e^{-\frac{1}{2}m}\chi^{m-1}\left(1+\dfrac{\chi^2}{n-m}\right)^{\frac{1}{2}n}\left(1+\dfrac{\chi^2}{n}\right)^{\frac{1}{2}}}{(n-m)^{\frac{1}{2}(n-1)}\,2^{\frac{1}{2}m-1}(\frac{1}{2}m-1)!}. \tag{25}$$

To this (20) will apply a correcting factor

$$\left(1-\frac{1}{2nk^2S\alpha^2}\right)^{\frac{1}{2}(m-1)} = \left(1-\frac{1}{\chi^2}\right)^{\frac{1}{2}(m-1)}, \tag{26}$$

and (21) a factor

$$\left(1-\frac{m-3}{\chi^2}\right)^{\frac{1}{2}}. \tag{27}$$

We can simplify (25) a little, on the hypothesis that m/n is small, by expanding $(n-m)^{\frac{1}{2}(n-1)}$ in powers of m/n and neglecting the difference between χ^2/n and $\chi^2/(n-m)$. Then we have

$$K = \frac{n^{\frac{1}{2}}\exp\left(\frac{1}{4}\frac{m^2}{n}-\frac{m}{2n}\right)}{2^{\frac{1}{2}m-1}(\frac{1}{2}m-1)!}\,\chi^{m-1}\left(1+\frac{\chi^2}{n-m}\right)^{-\frac{1}{2}(n-3)}, \tag{28}$$

with the extra factors from (26) and (27) The change in the form of the last factor from the simple $\exp(-\frac{1}{2}\chi^2)$ of the problem of chance is due, of course, to the fact that the standard error here is initially unknown, and corresponds to the difference between the normal law of error and "Student's" distribution. My test for series of measures (1937 *a*) needs an analogous change when the numbers of observations are not very large.

It has been supposed up to this point that the functions under consideration arise at the beginning of the investigation. Usually, however, a number of other parameters, the significance of which is not in doubt, must also be found from the data. The effect of this is to increase the uncertainties of the new ones, and this should be taken into account. If there are r of them, r new exponential factors appear in (5) and (6), and disappear when we integrate with respect to them, factors k^r being removed from (16) and (17) in the process. Thus our result must be divided by $(\sigma'/\sigma)^r$. The effect is that the index of the last factor in (28) must be reduced from $\frac{1}{2}(n-3)$ to $\frac{1}{2}(n-r-3)$. Thus if the analysis of a new set of data involves the determination of a large number of parameters already known to be relevant, the last factor, for a given set of residuals, may be substantially increased, and with it the difficulty of establishing any new ones In this formula, as adapted, the same definition of χ^2 by (24) is supposed retained.

If χ^2 is not much larger than m, even if it is significant, it will be impossible for many of the estimates of the parameters to exceed their standard errors. We may find that there is a significant departure from q in a very vaguely determined direction if m is large.

The foregoing analysis deals with the case where the functions introduced, as far as we know originally, might account for almost the whole of the outstanding variation. In many cases, particularly in astronomy, the previous information is enough to indicate the approximate limits of the outstanding variation and of the amplitudes of the terms to be tested, and the latter are appreciably smaller. In observations of the variation of latitude, for instance, I believe that the whole range of the effect is already well known to be under a quarter of the standard error of one observation, and is determinable only by combining numerous observations made under such conditions that it is practically certain that the bulk of the systematic errors will cancel. Chapman's determinations of the lunar atmospheric tide are a still more striking case. In such a problem 2 (7) does not represent the previous information, and must be replaced by

$$P(dc \mid \sim q, s, h) = dc/p,$$

where p is the permitted range of c indicated by previous considerations for the particular problem. The result is that in the above analysis ds/s^2 must be replaced by ds/ps wherever it occurs, and the range of integration for the

a_l is through a sphere of radius p instead of s. If the $\sigma(\alpha_l)$ are larger than p, the integration removes from $P(\sim q \mid \theta h)$ all factors depending on the α_l, and finally leaves K nearly 1. Thus until n is of order σ^2/p^2 the test is quite indecisive. This is what we should expect, but it is satisfactory that the analysis should lead to it. For much larger values of n the previous approximations are valid, and the result will be that $P(\sim q \mid h)$ is multiplied by σ/p and K by p/σ As the latter is small it becomes possible to infer a small systematic departure, when it is already expected to be small, more readily than when it is not.

6. *Test for independence from means of measures.* It is often found that summary values found by statistical treatment do not agree as well as the usual theory of combination of observations would imply. Now one possible source of error in this theory is that it assumes the errors of all observations independent. This is the type of assumption that leads to applications of Bernoulli's theorem, and when this happens I think that we should always look out for danger and test the hypothesis in question as soon as sufficient material is available. The common statement that the standard error of a mean is the standard error of one observation divided by the square root of the number of observations is not justified by the theory of probability, as is often stated; it is the result of the theory combined with the hypothesis of independence of the errors, and the latter is open to doubt. There is evidence, for instance, that the personal equation of an observer varies from time to time, so that there is a systematic effect running through a series of observations and therefore violating the hypothesis. We need a test that will reveal such an effect if it exists

Suppose that we have mr observations arranged in order, and divided into m groups of r each. The whole mr yield a mean $\bar{x}$ and a standard deviation σ The lth group alone gives mean $\bar{x}_l$ and standard deviation σ_l. If we take the mean of the group means we recover $\bar{x}$, and their standard deviation about it is τ The question is whether τ is consistent with the value $\sigma/\sqrt{r}$ inferred from the usual theory, or whether it is sufficiently greater to indicate that besides the random errors there is a systematic effect that may affect the whole of a range and possibly be reversed in the next.

We shall regard $\bar{x}$ as an estimate of a true value a. The hypothesis of complete independence is our q, and we shall suppose that the probability of one observation x, on this hypothesis, is given by

$$P(dx \mid q, a, s, h) = \frac{1}{\sqrt{(2\pi)}\,s}\exp\left\{-\frac{(x-a)^2}{2s^2}\right\}dx. \tag{1}$$

On the comparison hypothesis $\sim q$ the probability of an observation is distributed normally about a value a_l special to the group. Then

$$P(dx \mid \sim q, a_l, s', h) = \frac{1}{\sqrt{(2\pi)}\, s'} \exp\left\{-\frac{(x-a_l)^2}{2s'^2}\right\} dx. \tag{2}$$

We shall suppose also that the probabilities of the a_l's are distributed normally about a with standard error t, so that

$$P(da_l \mid \sim q, a, s', t, h) = \frac{1}{\sqrt{(2\pi)}\, t} \exp\left\{-\frac{(a_l-a)^2}{2t^2}\right\} da_l. \tag{3}$$

Since in either case s^2 is the expectation of the square of the departure of an observation from a, we shall have

$$s^2 = s'^2 + t^2 \tag{4}$$

We treat s as originally unknown and capable of any positive value, so that

$$P(ds \mid qh) = P(ds \mid \sim qh) \propto ds/s, \tag{5}$$

and t as unknown and capable of accounting for any fraction of s, so that

$$P(dt \mid \sim q, s, h) = dt/s. \tag{6}$$

Combining these probabilities by the product rule we have

$$P(q, ds\,da \mid h) \propto da\,ds/s, \tag{7}$$

$$P(\sim q, ds\,da\,dt\,da_1 \ldots da_m \mid h) \propto \frac{da\,ds\,dt}{s^2}\left(\frac{1}{2\pi t^2}\right)^{\frac{1}{2}m} \exp\left\{-S\frac{(a_l-a)^2}{2t^2}\right\} \Pi(da_l). \tag{8}$$

The probabilities of the data, given the various parameters, are proportional to

$$P(\theta \mid q\,ds\,da\,h) \propto s^{-mr} \exp\left[-\frac{mr}{2s^2}\{(a-\bar{x})^2+\sigma^2\}\right], \tag{9}$$

$$P(\theta \mid \sim q\,ds\,da\,dt\,da_1 \ldots da_m h) \propto s'^{-mr} \exp\left[-\frac{r}{2s'^2} S\{(a_l-\bar{x}_l)^2+\sigma_l'^2\}\right]. \tag{10}$$

Then by the principle of inverse probability

$$P(q\,ds\,da \mid \theta h) \propto da\frac{ds}{s} s^{-mr} \exp\left[-\frac{mr}{2s^2}\{(a-\bar{x})^2+\sigma^2\}\right], \tag{11}$$

$$\begin{aligned} P(\sim q\,ds\,da\,dt\,da_1 \ldots da_m \mid \theta h) \\ \propto \frac{da\,ds\,dt}{s^2}\left(\frac{1}{2\pi t^2}\right)^{\frac{1}{2}m} \exp\left[-S\frac{(a_l-a)^2}{2t^2}\right] \Pi\, da_l \\ \times s'^{-mr} \exp\left[-\frac{r}{2s'^2} S\{(a_l-\bar{x}_l)^2+\sigma_l^2\}\right]. \end{aligned} \tag{12}$$

We find

$$\frac{(a_l-a)^2}{t^2}+\frac{r(a_l-\bar{x}_l)^2}{s'^2}=\left(\frac{1}{t^2}+\frac{r}{s^2}\right)\left[a_l-\left(\frac{a}{t^2}+\frac{r\bar{x}_l}{s'^2}\right)\Big/\sqrt{\left(\frac{1}{t^2}+\frac{r}{s'^2}\right)}\right]^2+\frac{r(a-\bar{x}_l)^2}{s'^2+rt^2}. \quad (13)$$

Hence by integration with regard to the a_l,

$$P(\sim q\,ds\,da\,dt\,|\,\theta h)\propto\frac{da\,ds\,dt}{s^2}\,\frac{s'^{-mr+m}}{(s'^2+rt^2)^{\frac{1}{2}m}}\exp\left[-S\frac{r\sigma_l^2}{2s'^2}-\frac{r}{2(s'^2+rt^2)}S(a-\bar{x}_l)^2\right] \quad (14)$$

We have the identities

$$S(a-\bar{x}_l)^2=m(a-\bar{x})^2+m\tau^2, \quad (15)$$

$$Sr\sigma_l^2=mr(\sigma^2-\tau^2), \quad (16)$$

and integration with regard to a gives

$$P(q\,ds\,|\,\theta h)\propto s^{-mr}\exp-\frac{mr\sigma^2}{2s^2}\,ds, \quad (17)$$

$$P(\sim q\,ds\,dt\,|\,\theta h)\propto\frac{ds\,dt}{s^2}\,\frac{s'^{-mr+m}}{(s'^2+rt^2)^{\frac{1}{2}(m-1)}}\exp\left\{-\frac{rm\tau^2}{2(s'^2+rt^2)}-\frac{mr(\sigma^2-\tau^2)}{2s'^2}\right\}, \quad (18)$$

which becomes, on putting $t=su$,

$$\propto ds\,du\,\frac{(1-u^2)^{-\frac{1}{2}(mr-m)}s^{-mr}}{\{1+(r-1)u^2\}^{\frac{1}{2}(m-1)}}\exp\left[-\frac{mr(\sigma^2-\tau^2)}{2s^2(1-u^2)}-\frac{mr\tau^2}{2s^2\{1+(r-1)u^2\}}\right]. \quad (19)$$

Lastly, integrating with regard to s, we have

$$P(q\,|\,\theta h)\propto\sigma^{-(mr-1)}, \quad (20)$$

$$P(\sim q,du\,|\,\theta h)\propto du\,\frac{(1-u^2)^{\frac{1}{2}(mr-m)}}{\{1+(r-1)u^2\}^{\frac{1}{2}(m-1)}}\left\{\frac{\sigma^2-\tau^2}{1-u^2}+\frac{\tau^2}{1+(r-1)u^2}\right\}^{-\frac{1}{2}(mr-1)}, \quad (21)$$

and, if $\tau/\sigma=\gamma$,

$$\frac{1}{K}=\frac{P(\sim q\,|\,\theta h)}{P(q\,|\,\theta h)}=\int_0^1\frac{(1-u^2)^{\frac{1}{2}m-\frac{1}{2}}\{1+(r-1)u^2\}^{\frac{1}{2}(mr-1)}}{\{1+(r-\gamma^2r-1)u^2\}^{\frac{1}{2}(mr-1)}}\,du \quad (22)$$

This expression is exact.

Three checks are possible at this stage. If $m=1$, the observations are all lumped in one group and there is no information for testing a variation in different groups. In this case τ^2 and γ^2 are zero. On substitution in (22) we find $K=1$, so that the test makes no decision—which is the correct decision.

If $r=1$, all the observations are in different groups and there is no means of separating a systematic variation between groups from the variation of individual observations. In this case $\tau^2 = \sigma^2$, $\gamma^2 = 1$, and on substitution we find again $K = 1$.

If $m = 2$, the test should reduce to the test for comparing the means of two series of observations for a systematic difference. This test was, for large numbers of observations,

$$K = \left(\frac{2}{\pi}\frac{\sigma^2+\tau^2}{\sigma^2/m+\tau^2/n}\right)^{\frac{1}{2}} \exp\left\{-\frac{(\bar{y}-\bar{x})^2}{2\sigma^2/m+2\tau^2/n}\right\}, \tag{23}$$

where $\bar{x}$ and $\bar{y}$ are the means, m and n the numbers of observations, and σ and τ the standard deviations in the two series. Here we must replace m and n by r, σ and τ by σ, and $\bar{y}-\bar{x}$ by 2τ. Then K reduces to

$$K = \left(\frac{2r}{\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{r\tau^2}{\sigma^2}\right) = \left(\frac{2r}{\pi}\right)^{\frac{1}{2}} \exp(-r\gamma^2). \tag{24}$$

Putting $\gamma = 0$ in (22) we easily verify the first factor. When γ is not zero the behaviour of the integrand is complicated, as we shall see in the general case.

In general, if $\gamma = 0$, we find approximately

$$K = \{2(m-1)\,r/\pi\}^{\frac{1}{2}}, \tag{25}$$

so that the support for q, in the most favourable case, is of order $n^{\frac{1}{2}}$, as usual. This experimental result would however be very unlikely to occur in such a problem.

The large indices make manipulation troublesome, but we may substitute

$$w^2 = \frac{ru^2}{1+(r-1)\,u^2}. \tag{26}$$

Then

$$\frac{1}{K} = \int_0^1 \frac{r(1-w^2)^{\frac{1}{2}(m-1)}\,dw}{(1-\gamma^2w^2)^{\frac{1}{2}(mr-1)}\{r-(r-1)\,w^2\}^{\frac{1}{2}}}. \tag{27}$$

For $\gamma^2 < 1/r$, nearly, the integrand decreases steadily through the range, and the integral is, approximately,

$$\frac{1}{K} = \left\{\frac{\pi}{2mr(1-r\gamma^2)}\right\}^{\frac{1}{2}}. \tag{28}$$

This is valid so long as $1-r\gamma^2$ is greater than a quantity of order $m^{-\frac{1}{2}}$.

For $\gamma^2 = 1/r$, the factors with high exponents behave like $\exp(-\frac{1}{4}mw^4)$. Neglecting higher powers we obtain the approximation

$$\frac{1}{K} = \frac{2^{\frac{1}{2}}\,.\,\frac{1}{4}!}{r^{\frac{1}{4}}m^{\frac{1}{4}}} = \frac{1{\cdot}28}{r^{\frac{1}{4}}m^{\frac{1}{4}}}. \tag{29}$$

There continues to be support for independence up to the value of γ^2 that corresponds to the expectation of τ^2 given q. This was to be expected, but the change in the index of m is a new feature. It comes from the vanishing of the terms in w^2 near this transition value of γ^2. Epistemologically, it comes from the fact that if t is a small quantity of the first order it makes only a second-order change in the expectation of τ, and therefore of γ. In our other problems a first-order change in a quantity to be determined implies a first-order change in the observational quantity to be used to estimate it. The result is that a value of γ^2 near to $1/r$ corresponds to a larger range of possible values of t, if t is not zero, than it would for the usual linear relation, and the coincidence required on hypothesis $\sim q$ is less remarkable. (It was, however, the fact that a small real amplitude makes a second-order change in χ^2 that produced the extra factors in the problem of chance, though they were not important enough to alter the index.)

For $\gamma^2 > 1/r$ the integrand increases from the origin to a maximum and declines again to zero. The maximum is near

$$w^2 = \frac{r}{r-1}\left(1-\frac{1}{r\gamma^2}\right), \tag{30}$$

and if we denote this value of w by w_0 the integrand is proportional to

$$\exp-\frac{m(r\gamma^2-1)(r-1)^2\gamma^2}{r(1-\gamma^2)^2}(w-w_0)^2, \tag{31}$$

and we can approximate to the integral by the method of steepest descents, provided that

$$r\gamma^2-1 > (1/m)^{\frac{1}{2}}. \tag{32}$$

We find $$K = \frac{r^{\frac{1}{2}mr-2}m^{\frac{1}{2}}}{(r-1)^{\frac{1}{2}mr-\frac{1}{2}m-1}\pi^{\frac{1}{2}}}(r\gamma^2-1)^{\frac{1}{2}}\gamma^{m-3}(1-\gamma^2)^{\frac{1}{2}mr-\frac{1}{2}m-1} \tag{33}$$

This can be simplified by putting

$$\gamma^2 = \frac{1}{r}+\frac{\beta}{r\sqrt{m}} \tag{34}$$

and expanding to order β^2. We find finally

$$K = \frac{r^{\frac{1}{2}}m^{\frac{1}{4}}}{\pi^{\frac{1}{2}}}\beta^{\frac{1}{2}}\exp(-\tfrac{1}{4}\beta^2). \tag{35}$$

Comparing with (29) and (32) we see that this should be of the right order of magnitude for $\beta = 1$ and a good approximation at higher values.

This problem is one of the first treated by Fisher's z-distribution (1924, 1928) and since the factor in K that depends on the observed values has so far usually been found to agree closely with those found by direct methods it is of interest to make a corresponding comparison here. (An exception has, however, been noticed above for measures when the normal equations are highly non-orthogonal.) In this problem we could regard $\sigma^2-\tau^2$ and τ^2 as giving two independent estimates of s^2 if q was true. These estimates would be

$$s_1^2 = \frac{mr}{m-1}\tau^2, \text{ based on } m-1 \text{ degrees of freedom;} \tag{36}$$

$$s_2^2 = \frac{r(\sigma^2-\tau^2)}{r-1}, \text{ based on } m(r-1) \text{ degrees of freedom.} \tag{37}$$

Fisher defines

$$2z = \log\frac{s_1^2}{s_2^2}, \tag{38}$$

and gives a form, on the hypothesis that all deviations are completely independent, for the probability density of z. This, in my notation, is

$$P(dz\,|\,qh) \propto \frac{e^{(m-1)z}}{\{(m-1)e^{2z}+m(r-1)\}^{\frac{1}{2}(mr-1)}}\,dz. \tag{39}$$

Substituting for z in terms of σ^2 and τ^2, we have

$$e^{2z} = \frac{m(r-1)}{m-1}\frac{\gamma^2}{1-\gamma^2} \tag{40}$$

and

$$P(d\gamma\,|\,qh) \propto \gamma^{m-2}(1-\gamma^2)^{\frac{1}{2}mr-\frac{1}{2}m-1}\,d\gamma \tag{41}$$

This form resembles (33) very closely, even to the power of $1-\gamma^2$. One of the factors γ in it, however, is replaced in my solution by $(\gamma^2-1/r)^{\frac{1}{2}}$ on account of the unusual nature of the effect of t on the quantity used to estimate it. It is this factor that gives the $\beta^{\frac{1}{2}}$ in (35), the variation of which is of secondary importance in comparison with that of the exponential factor. Considering the difference in the methods of approach the agreement is very satisfactory.

It is sometimes said that too small a value of χ^2, or in this case of τ^2, gives as good reason for rejecting the hypothesis to be tested as too large a value does. In my solutions the support for the hypothesis q always increases right up to χ^2 or $\tau^2 = 0$. This appears to be because I have not considered, among the alternative hypotheses denoted by $\sim q$, the types that could lead to spurious agreements. For instance, a negative correlation between consecutive observations in a series would make the means by ranges vary less than would be expected on the hypothesis of independence. This might happen if an observer tended to compensate an error in one direction by making the next one in the opposite direction. Special discussion would be needed to test a hypothesis of this type I do not think that instances of it are common; much commoner is a positive correlation that leads to an underestimate of the uncertainty and hence to too large a χ^2 when different series are compared.

We notice from (14) and (15) that if $\sim q$ is established by the test the posterior probability of a is distributed about $\bar{x}$ with standard error given by

$$\sigma_a^2 = \frac{s'^2 + rt^2}{rm}. \tag{42}$$

But, apart from sampling errors, the maximum of (18) gives

$$s'^2 + rt^2 = \frac{rm\tau^2}{m-1}, \tag{43}$$

so that

$$\sigma_a^2 = \tau^2/(m-1) \tag{44}$$

Hence if a test of this type shows that there are systematic errors affecting ranges the precision of the mean should not be found from the original scatter of the observations, but by treating the means for the ranges as m independent measures, and using their scatter as the datum for estimating the uncertainty of a. This does not exclude the possibility that closer investigation may lead to the discovery of further correlations between even these; but the range means will determine a minimum uncertainty if the hypothesis of independence has to be rejected. This method is often used in practice; but unfortunately it is also often used when the means are more accordant than their standard errors would suggest, and then leads to an underestimate of the uncertainty.

7. *Numerical examples.*

7·1. *Periodicity in measures. the node of Venus.* Dr H. S. Jones (1929) gives the following table of discrepancies between the observed and calculated secular changes of the orbits of the inner planets, after allowing for Einstein's correction to the perihelia and for corrections to the masses made from independent evidence The unit is 1″ per century:

	de	$e\,d\varpi$	di	$\sin i\,d\Omega$
Mercury	−0·90 ± 0·50	−0·89 ± 0·40	+0·38 ± 0·80	+0·85 ± 0·45
Venus	+0·17 ± 0·20	−0·13 ± 0·20	+0·40 ± 0·30	+0·76 ± 0·12
Earth	+0·01 ± 0·09	−0·04 ± 0·12	−0·06 ± 0·10	—
Mars	+0·29 ± 0·20	+0·42 ± 0·25	−0·31 ± 0·16	+0·14 ± 0·16

The respective contributions to χ^2 are:

	de	$ed\varpi$	di	$\sin i\,d\Omega$	Total
Mercury	3·24	4·95	0·23	3·56	11·98
Venus	0·72	0·42	1·78	40·1	43·02
Earth	0·01	0·11	0·36		0·48
Mars	2·11	2·82	3·74	0·77	9·44
Total	6·08	8·30	6·11	44·43	64·92

Such a χ^2 on 15 degrees of freedom is overwhelming. The diagram given by Yule and Kendall (1937) gives the 0·01 % limit at $\chi^2 = 44$. But 40·1 of it comes from the node of Venus, which is viewed with astonishing equanimity by those writers on relativity who describe the agreement as completely satisfactory. If it is omitted, χ^2 for 14 degrees of freedom sinks to 24·8, which would make the P integral equal to 0·03, and might just possibly be due to random error.

On the present theory, we must regard these residuals as parameters arising in pairs, with the exception of the change of the obliquity of the ecliptic, the companion of which would be the constant of precession and is taken as a datum. Thus de and $ed\varpi$ can be regarded as coefficients of $t \cos nt$ and $t \sin nt$ in the longitude, di and $\sin i\,d\Omega$ as coefficients in the latitude. The reduction from right ascension and declination, with respect to the earth, to celestial latitude and longitude with regard to a fixed set of axes in the sky, may be regarded as a secondary complication. In testing the node of Venus for significance we must therefore take it with its companion, the change of inclination For these two we have, therefore,

$$n = 12319; \; m = 2; \; \chi^2 = 41\cdot9$$

With this number of observations we can replace the last factor in 5 (28) by $\exp(-\tfrac{1}{2}\chi^2)$; then

$$K = 111\sqrt{(42)}\,e^{-21\cdot0} = 5\cdot4 \times 10^{-7}.$$

This does not allow for the fact that we have selected this pair as an extreme departure out of 8 To allow for this we must multiply by $8/\log_e 2 = 11\cdot5$, giving $6\cdot2 \times 10^{-6}$. The odds are therefore about 160,000 to 1 that this is a genuine departure, whatever its explanation may be.

In a previous discussion (1936, p. 445) I found that, with so large a number of observations, the residuals would be just within the range attributable to random error. The change arises from three sources. In my previous test the coefficient of $\exp(-\tfrac{1}{2}\chi^2)$ was 6160, which has been divided by about 9 by the correction of the test. This change is roughly balanced by the allowance for selection. The chief changes are due to Jones's use of a corrected mass of the earth, in comparison with that of the sun, to agree with more accurate determinations of the solar parallax than were available to Newcomb, this has increased the residual of the node. Newcomb's standard errors included those of the masses, which was an undesirable procedure since it implied that the uncertainties were not independent. Jones has omitted these (they may now be considered negligible) and the standard errors are correspondingly reduced. Thus this pair of contributions to χ^2 have risen from 13·9 to 41·9

It may be remarked that the observations covered about 140 years and that the observable effect in this time would be about 1″. The diameter of the disk of Venus at maximum elongation is 26″. It may, perhaps, be doubted whether the bisection of an oddly shaped object like the visible crescent of Venus could be trusted to this accuracy without systematic error, or even whether the systematic error could be trusted to remain the same for astronomers over 140 years In making this suggestion, however, I am opposing the authority of Newcomb, who considered the node remarkably free from systematic error.

Apart from this pair, the largest contribution to χ^2 comes from de and $ed\varpi$ for Mercury, and is 8·2. The factors depending on χ^2 would therefore be $2{\cdot}9\,e^{-4\cdot1} = 1/21$ roughly, and would be overcome by the factor $n^{\frac{1}{2}}$ if there were more than 441 observations. As there are several thousands there is no reason to regard these contributions as significant, though the general magnitude of the contributions from Mercury and Mars suggests that the standard errors may have been slightly underestimated.

7·2. *Test for independence of errors Pearson's data.* Karl Pearson (1902) has given the results of six series of observations designed to test the constancy of the personal equation and the correlation of the errors of different observers. The first test consisted of the bisection by eye of a line, which was afterwards measured The second was essentially a time observation of a moving bright line. There were three observers for each type of experiment. Each observer made 500 bisection observations and 519 bright-line observations. Pearson tabulates the means of groups of successive observations for each individual. For the bisection series there are 20 groups of 25 each; for the bright-line series there are 16 groups of 27 to 37 each and an odd group of 17. To preserve symmetry I have ignored the last group and neglected the difference between 27 and 37. The data are then in a form suitable for the application of the test for independence by the consistency of the group means.

For the bisection series the results are:

$$r = 25;\ m = 20;\ r^{\frac{1}{2}}m^{\frac{1}{2}}/\sqrt{\pi} = 5{\cdot}96.$$

Observer	Mean personal equation	$10^4\sigma^2$	$10^4\tau^2$	γ^2
1	+0·01235	6·026	0·561	0·0931
2	−0·00477	9·397	3·404	0·3623
3	−0·00469	6·892	3·794	0·5505

All the values of γ^2 are larger than the expectation on the hypothesis of randomness, which would be 0·0400. To apply the test we first compute $\beta = (r\gamma^2 - 1)\sqrt{m}$; it is interesting to show the various factors of K separately:

Observer	$r^{\frac{1}{2}}m^{\frac{1}{2}}/\sqrt{\pi}$	β	$\beta^{\frac{1}{2}}$	$e^{\frac{1}{2}\beta^2}$	K
1	5·96	5·937	2·436	6700	0·00216
2	5·96	36·035	6·002	10^{141}	4×10^{140}
3	5·96	57·07	7·555	10^{352}	4×10^{-352}

For the bright-line series we take mean values for the numbers in the groups, namely $r = 502/16 = 31{\cdot}4$, $m = 16$ The mean taken is an unweighted one for the groups retained.

Observer	Mean personal equation	σ^2	τ^2	γ^2
1	+0·1828	1·414	0·0939	0·0664
2	−1·1617	1·376	0·1377	0·1001
3	−0·5373	3·318	0·4628	0·1395

Observer	$r^{\frac{1}{2}}m^{\frac{1}{2}}/\sqrt{\pi}$	β	$\beta^{\frac{1}{2}}$	$e^{\frac{1}{2}\beta^2}$	K
1	6·32	4·464	2·113	138·6	0·0963
2	6·32	8·572	2·928	$9{\cdot}42 \times 10^7$	2×10^{-6}
3	6·32	13·520	3·677	7×10^{19}	3×10^{-18}

The results are very striking. Not one of the six series gives a value of K that supports the hypothesis that the errors are independent, and four are overwhelmingly against it. The correlation between neighbouring observations is such that the mean of 25 consecutive observations, in the bisection experiments, is no more accurate than that of 2 to 11 independent ones should be; and in the bright line ones, the mean of 31 is no more accurate than that of 7 to 15 independent ones.

Pearson does not summarize his means by ranges except in the form of graphs, on which the expected random variation $\sigma/\sqrt{r}$ is not marked, but his conclusions are substantially the same as these. It is clear that the hypothesis of independence of the errors can give only a minimum uncertainty, and that when a large number of observations are combined additional checks should be applied to test it and allow for its failure, otherwise an appearance of accuracy may be obtained that is entirely spurious.

With these numbers of observations K would be 1 about $\beta = 3$, so that γ^2 would differ from $1/r$ by $3/rm^{\frac{1}{2}}$. Then for the difference to be just significant

$$\tau = \sigma\sqrt{\frac{1}{r}}\left(1+\frac{3}{2\sqrt{m}}\right).$$

But the standard error usually given for a standard error based on m observations is $(2m)^{-\frac{1}{2}}$ of the latter. We could regard τ as such an estimate, and if it differs from $\sigma r^{-\frac{1}{2}}$ by $3/\sqrt{2} = 2{\cdot}12$ times its estimated standard error we could regard the difference as genuine. It is curious how the rough rule that with ordinary numbers of observations a departure of about twice the standard error is just significant persists in these tests.

This lack of independence is relevant to the question of the value of special studies as against the use of reports from the observing stations in the construction of seismological time tables. The former are often recommended on account of the reduced error of one observation. In some cases this is genuine, though the difference is not great, but when the same observer reads the whole of the records there is a serious danger that the apparent smoothness is achieved by a correlation between consecutive errors and not by any real increase of accuracy. I have a strong suspicion that this is particularly true of what is called a "careful observer", who may read the same record again and again until he is "sure", and thereby, quite unconsciously, allow his readings to be influenced by those of other records. The true accuracy, in such a case, is better determined by the first readings than by the consistency of the final ones.* When the readings for each earthquake are all made by different observers there is much more prospect of achieving independence of the errors. I am confirmed in this opinion by the fact that, although the law of error deviates widely from the normal in these conditions, when allowance is made for this deviation the comparison of different series for the phases regularly observed reveals no unexplained discrepancies, even though the standard errors have in most cases been reduced by combining observations to about a sixth or less of that of one observation. It is true that many of the alleged discrepancies between the results of special studies and those based on reports are spurious, since hardly any of them are associated with any estimate of the standard error at all, and some are due to graphical methods, errors in arithmetic, interpolation over long ranges where there are hardly any observations to serve as a check, or to the inclusion of observations affected by known types of systematic error, such as, especially, late readings due to smallness of the movement. But this is not the present point, which is that errors in the reports satisfy the condition of independence, and that consequently the accuracy apparently attained is genuine, while those in special studies are open to suspicion until their apparent accuracy is confirmed by comparison with other series The proper function of special studies still seems to me to be, as Bullen and I remarked in our original paper (1935), to test doubtful points such as the separation of difficult phases, where reports lead to indecisive results, with due regard to the risk of the observer finding what he expects to find—which certainly varies very much from one observer to another. The computation of χ^2 in comparing series serves two purposes, since χ^2 would be increased either by a genuine systematic difference or by an underestimate of the standard errors When a normal value is found, therefore, it confirms both the absence of a systematic difference and the independence of the errors I have in fact made considerable use of special studies in treating the more difficult phases, but this test has always been applied and has usually, but not always, shown that the series are comparable I do not wish, therefore, to disparage special studies, but merely to point out that such a check is always necessary before their apparent accuracy is accepted.

---

* Or independence could be achieved by reading the records in a random order given by card drawing or Tippett's numbers.

8. *The combination of tests* It sometimes happens that a series of estimates of a parameter consistently give values with the same sign and running up to about twice their standard errors. None of them taken by itself would be regarded as significant, but when they all say the same thing one begins to wonder whether there may not be something to be said for them after all. A treatment is suggested by the problem of sampling to test an even chance. The appropriate formula is

$$K = (2n/\pi)^{\frac{1}{2}} \exp(-\tfrac{1}{2}\chi^2). \tag{1}$$

Now suppose that we have a sample of 1000 and that the departure makes K less than 1 If we divide the data into 9 groups we divide the outside factor by 3; but at the same time we multiply all the standard errors by 3 and divide the contribution to χ^2 from a genuine departure by 9. Thus a departure that would be shown by a sample of 1000 may not be shown by any one of its sections. Since K is the factor to be applied to the ratio $P(q \mid h)/P(\sim q \mid h)$ to give the posterior probabilities and all the separate $K_1, K_2, \ldots, K_9$ may be more than 1, and yet the K given by taking the whole sample together is less than 1, it appears that we have an inconsistency. This arises from an insufficient analysis of the alternative $\sim q$. The hypothesis q is a definitely stated proposition, leading to definite inferences. $\sim q$ is not, because it contains an unknown parameter,* which we may denote by p, and would be $\frac{1}{2}$ on q, but may be anything from 0 to 1 on $\sim q$. Anything that alters the range permitted to p will alter the inferences given by $\sim q$. Now the first sub-sample does alter this range. We may start with probability $\frac{1}{2}$ concentrated at $p = \frac{1}{2}$ and the other $\frac{1}{2}$ spread from 0 to 1 In general the data of the first sample will alter the ratio of these amounts and may increase the probability that p is 0; but it also greatly changes the distribution of the probabilities of p given $\sim q$, which are now nearly normally distributed about the sampling ratio with an assigned standard error. It is from this state of things that we start when we make our second subsample, not from a uniform distribution of the probability of p, supposing that it is not $\frac{1}{2}$. The permitted range has been cut down to something of the order of the standard error of the sampling ratio given by the first sample. Consequently the outside factor in (1) is greatly reduced and the second sample may support $\sim q$ at a much smaller value of $p-\frac{1}{2}$ than would be the case if it started from scratch. We cannot therefore combine tests by simply multiplying the values of K.

A general argument shows what the result must be. If p_1 and p_2 are two hypotheses, and we have two sets of data θ_1 and θ_2, the original information being h,

$$\frac{P(p_1 \mid \theta_1 h)}{P(p_1 \mid h)\,P(\theta_1 \mid p_1 h)} = \frac{P(p_2 \mid \theta_1 h)}{P(p_2 \mid h)\,P(\theta_1 \mid p_2 h)}, \tag{2}$$

$$\frac{P(p_1 \mid \theta_1\theta_2 h)}{P(p_1 \mid \theta_1 h)\,P(\theta_2 \mid p_1\theta_1 h)} = \frac{P(p_2 \mid \theta_1\theta_2 h)}{P(p_2 \mid \theta_1 h)\,P(\theta_2 \mid p_2\theta_1 h)}, \tag{3}$$

by two applications of the principle of inverse probability By multiplication,

$$\frac{P(p_1 \mid \theta_1\theta_2 h)}{P(p_1 \mid h)\,P(\theta_1 \mid p_1 h)\,P(\theta_2 \mid p_1\theta_1 h)} = \frac{P(p_2 \mid \theta_1\theta_2 h)}{P(p_2 \mid h)\,P(\theta_1 \mid p_2 h)\,P(\theta_2 \mid p_2\theta_1 h)}. \tag{4}$$

But

$$P(\theta_1 \mid p_1 h)\,P(\theta_2 \mid p_1\theta_1 h) = P(\theta_1\theta_2 \mid p_1 h), \tag{5}$$

$$P(\theta_1 \mid p_2 h)\,P(\theta_2 \mid p_2\theta_1 h) = P(\theta_1\theta_2 \mid p_2 h), \tag{6}$$

and therefore

$$\frac{P(p_1 \mid \theta_1\theta_2 h)}{P(p_1 \mid h)\,P(\theta_1\theta_2 \mid p_1 h)} = \frac{P(p_2 \mid \theta_1\theta_2 h)}{P(p_2 \mid h)\,P(\theta_1\theta_2 \mid p_2 h)}, \tag{7}$$

which is the result of applying the principle of inverse probability to the data θ_1 and θ_2 simultaneously. Thus it does not matter in what order we introduce our data; as long as we start with the same data and finish with the same additional data, the final results will be the same. The principle of inverse probability cannot lead to inconsistencies. To apply it in the wrong way to the data available may, and often does, and in this case the wrong way is to ignore θ_1 in (3), as if we started in both stages from previous ignorance, which may be true for the first but cannot possibly be true for the second.

---

* This distinction appears also in Fisher's theory (1935, p. 19).

Fortunately it is not necessary in this case to carry out the analysis in detail, because the totality of the data, which would be denoted by $\theta_1\theta_2 \ldots \theta_n = \theta$, is the complete sample, and we have the result for it already. It follows that the method of combining samples in a test is to add the values of n in the outside factor and to use a χ^2 based on the ratio of the deviation from $\frac{1}{2}$ of the sampling ratio based on all the samples together to its standard error.

Analogous considerations will apply to measures so long as the standard errors of one observation are equal in the data combined. If they differ considerably some modification may be needed, since we have seen that two departures with the same standard error may give different results in a test when the same standard errors of the estimates are based on different numbers of observations of different accuracies. The outside factor in such a case will not be obtained by simply adding the numbers of observations, since what it really depends on is the ratio of the range of the permitted variation to the standard error of the result The former is fixed by the smallest range indicated and therefore by the most accurate observations, and the less accurate ones have nothing to say about it. It is only when they are numerous enough to give a standard error of the result less than the range permitted by the more accurate ones that they have anything additional to say. If they satisfy this condition the outside factor will be got by taking s from the most accurate observations and α and its standard error σ from all the series together.

9. *The use of integrals in significance tests* Though the functions of the observed values that appear in these tests are usually identical, or nearly so, with those in the tests used by modern statisticians, there are some differences. Here they appear directly, in the usual forms only their integrals are used. In the frequency problem, for instance, the information supplied by the observations is summed up, approximately, in $\chi^{m-1}\exp(-\frac{1}{2}\chi^2)$, but the usual form of the test depends on the integral of this quantity, in the form

$$P(\chi^2) = \int_{\chi}^{\infty} \chi^{m-1}\exp(-\tfrac{1}{2}\chi^2)\,d\chi \Big/ \int_0^{\infty} \chi^{m-1}\exp(-\tfrac{1}{2}\chi^2)\,d\chi.$$

Further differences are that in my form m is to be taken as 1 unless there is specific reason to take more than one degree of freedom together, and even then is only the actual number of degrees so associated. In Pearson's form it is the whole number of groups. My χ^2 again is only the contribution to χ^2 given by the degrees of freedom being considered; Pearson's is the complete χ^2, including all the degrees of freedom, whether variation in them is random or systematic In Fisher's usage, χ^2 is given by the ratios of the amplitudes found by maximum likelihood to their standard errors, and his χ^2 and m are identical with mine where there is reason to take several degrees of freedom together, but he uses the integral P as a test, so that his standpoint is intermediate between Pearson's and mine In this way he avoids the great random variation of χ^2 taken over all groups that so often makes it impossible to detect a genuine systematic effect when the test is applied in Pearson's way, further, he in some cases recommends an allowance for selection in a way qualitatively similar to mine (Fisher 1935, pp 65–6).

The use of the integral goes back at least to Chauvenet's criterion for the rejection of observations This criterion, still sometimes used, considered the probability, given the normal law, that n observations should include at least one residual equal to or greater than the largest actually found If this was less than $\frac{1}{2}$ the observation was rejected. The difficulty that has struck many students is, why should the limit be taken at the largest? There was more to be said, apparently, for choosing the second largest. The probability of a residual just equal to the largest is necessarily infinitesimal until the observations are made, or at most of the order of magnitude of the ratio of the rounding-off error to the standard error. Consequently the integral giving the probability of a residual equal to or greater than the largest depends entirely on the contribution from larger deviations, which have by hypothesis not occurred The use of the criterion, as recommended by Chauvenet, means that an observation is rejected because observations that have not occurred were unlikely. One might indeed say that the fact that they have not occurred is confirmation of the hypothesis and that the observation should be retained. If the lower limit of the integral was taken at the second largest residual instead of the largest there would be at any rate one observation relevant to the test, but as actually taken the test is illusory.

Pearson's procedure (1900) was as follows. The probability, given a suggested distribution of chance, of the actual observations in the groups, is evaluated, and it is found that the factor that depends on the observations is $\exp(-\frac{1}{2}\chi^2)$, subject to certain approximations. It is thus shown that the likelihood is proportional to this function and that maximum likelihood and minimum χ^2 are equivalent; the information contained in the observations is entirely summed up in χ^2. Pearson, however, goes further, and proceeds to find the probability distribution for χ^2, given the same trial hypothesis as before, and gets the rule $\chi^{n-1}\exp(-\frac{1}{2}\chi^2)\,d\chi$. One may ask, why, when the value for the hypothesis and the actual observations is already known, should there be any interest in the probability of having got something else? But there was reason to be interested in the effect of different hypotheses on χ^2, keeping the observations the same. The additive property of χ^2, again, makes it possible to use it to separate the posterior probability distributions of the parameters that arise in varying the hypothesis; it could be used to estimate the ratios of all the possible variations of these to their standard errors and thus to test all separately. It is remarkable that Pearson, who repeatedly declared his support of the principle of inverse probability,* apparently did not notice this simple consequence of it, and continued to use the test in a way that lumped all degrees of freedom together. Finally he formed the P integral and used it to form a test of the correctness of the trial hypothesis, based on the probability, given that hypothesis, of a larger χ^2 than that observed, the actual χ^2 of course makes a negligible contribution, as in Chauvenet's criterion. A hypothesis that might be true is therefore rejected because it does not agree with observations that have not been made. The use of P in this sense is therefore a mistake.†

When the degrees of freedom are properly separated this is comparatively harmless Tests of significance are needed in any case, simply because even if the trial hypothesis was true, random errors would lead to determinations different from zero for a departure in every degree of freedom. If these were accepted as genuine we should expect to lose accuracy in prediction The question is to decide where to draw the line beyond which we should expect to gain accuracy by accepting the departures found. But as P is a monotonic function of χ^2, whatever the number of degrees of freedom, to fix a value of P and to fix one of χ^2 are the same thing. It would have been justifiable empirically to rely on the rough rule found by astronomers, that differences under twice the standard error usually tend to disappear with fuller data, while those over three times the standard error usually persist. Thus for one degree of freedom a contribution of 4 to χ^2 could have been taken as possibly genuine, one of 9 almost certainly so, and this would be on purely empirical grounds For any combination of associated degrees of freedom the distribution of χ is still roughly normal about a maximum, and the ratio of the actual ordinate to the maximum would have provided a test as good as the P integral gives. In one respect, indeed, it is better, since it is often pointed out that a very large value of P is a ground for suspicion. It is hard to justify this from P itself, but large values of P, like small ones, imply that the ordinate is small compared with the maximum. The kinds of inference drawn in such cases, however, would be different, usually too large a χ^2 would suggest a new parameter or a positive correlation between errors, too small a χ^2 would suggest a negative correlation, or that the data had been altered to suit a hypothesis

---

* The 1911 edition of *The Grammar of Science* contained the same excellent account of it as the earlier editions

† Yates (1934) has shown that the probabilities of small groups are better estimated if the limit for χ^2 is taken, not from the observed number in a group, but at $\frac{1}{2}$ more. In this way he made the observed value make a much larger contribution to P This procedure is recommended by Fisher (1937, p 97).

Any departure from the trial law, if found by least squares or maximum likelihood, would diminish χ^2 and therefore increase P; so long as the number of degrees of freedom is the same, therefore, if two precisely stated hypotheses are compared the one with the smaller P is the more probable, if there is no other reason for preferring either. But the ratio of the posterior probabilities is not that of the P's, but that of the $\exp(-\frac{1}{2}\chi^2)$, apart from a negligible difference in the outside factor. The situation is altered if we are discussing the introduction of a new adjustable parameter, because this can always be adjusted so as to increase P or reduce χ^2. An attempt to allow for this is sometimes made by reducing the degrees of freedom by one at the same time, but it is easy to see that if we do this and keep the reduction of P as the criterion we get unacceptable results. From Fisher's table we find that the 5 and 1 % levels for χ^2 for one degree of freedom are at 3·84 and 6·64. But the changes from 20 to 21 degrees of freedom at these levels are only 1·26 and 1·37 Thus if 20 degrees of freedom would be taken as random variation, an additional one contributing only 1·27 or 1·38 to χ^2 could lead to the assertion of a systematic difference, whereas three or five times these contributions would be needed if the random errors in the others were eliminated and it was tested directly This seems utterly contrary to common sense. It is interesting to notice that Fisher, by insisting on the separation of genuine possible departures from random error, without apparently being guided by the principle of inverse probability, arrived at a procedure closely resembling what it indicates, though stated in terms of P and not of the ordinate. I think, however, that the difference in premises is more apparent than genuine. I regard the theory of probability as a formal statement of common-sense reasoning, the principles of which are quite general and can be stated at the start; but even if this is not done an alert thinker may still notice the need for them in specific applications and thereby arrive at the same or very similar results.

It may be remarked that Fisher's z-distribution, for the analysis of variance, has been deliberately transformed so as to eliminate scale effects, and is nearly normal, and therefore the above relation between the ordinate and P will hold for it. For problems of estimation, where an effect is assumed to be present but is undetermined in magnitude, the integral has a definite place in the inverse theory, by the kind of argument that I used for "Student's" distribution (1937 *c*), it will give, for instance, the posterior probability that an estimated difference has the right sign. But if this was all that was meant by a significance test it would imply that the rejection of small differences implied a loss of accuracy in prediction. According to such a view astronomers, for instance, instead of finding the orbital elements and masses of the planets by least square solutions and predicting their places according to the law of gravitation, would do better to choose polynomials to fit all the observations exactly and extrapolate accordingly. Certainly no astronomer believes that, and I very much doubt whether anybody does, even of those who say that they reject the simplicity postulate.

10 *The excluded middle*. All inference from observation involves this alleged logical fallacy, which is a fallacy only so long as it is claimed that this inference is deductive I have shown its unimportance in relation to significance tests (1935, p 222), in which a new parameter is taken to be zero so long as the tests do not show that it is more probably something else. I have only noticed recently, however, that this could have been inferred from a theorem given by Dr Wrinch and me in 1921, and used then as the basis of the simplicity postulate It was shown (1937 *b*, p. 41) that if a general law p has a finite probability at any time, and gives a series of inferences $q_1, q_2, \ldots, q_n$, the effect of the verification of the successive inferences is to divide the probability of the law, in turn, by

$$P(q_1 \mid h),\ P(q_2 \mid q_1 h) \ldots P(q_n \mid q_1 \ldots q_{n-1} h)$$

Since none of these can exceed unity the probability of the law, given all the verifications, would become greater than 1 if they did not tend to 1. Hence so long as inferences from a law continue to be verified the probability of the next verification approaches certainty. But $P(q_n \mid q_1 q_2 \ldots q_{n-1} h)$ does not involve p at all, in other words there is a high probability that p will continue to be verified *whether it is true or not*. This means that if the "true" law (whatever that may mean) is something different from p, it must nevertheless be so like p as to have led to all the previous inferences from p, any law that led to different ones would have been discarded as in disagreement with observation. But if the true law has led to all the same inferences as p the presumption is that it will continue to do so, and therefore that p will continue to give the right inferences. This disposes of the "excluded middle" argument completely. Astronomers were not wrong in basing their predictions on Newton's law because the law turned out to be wrong, nor were geodesists wrong in using Euclid's theory because that is wrong too; their justification did not rest on the truth of the laws but on the fact that they had led to right predictions hitherto and might reasonably be expected to continue to do so The kind of caution that insists on stating every conclusion in the vaguest possible terms (except, possibly, the author's own) is not in accordance with scientific principles. The function of a general law is to make predictions that can be tested, even if at some unpredicted data in the future some of them may turn out to be wrong The more vaguely it is stated the less it fulfils this function. Even if the law should ever lead to wrong predictions its precise statement will have been the means of revealing the discrepancy and possibly leading to a better one, which would have to account also for all the previous verifications. It is usual to find some verbal concession to the "excluded middle" in scientific work, and this is quite undesirable so long as we regard mathematics as something to be used and not to be worshipped; it is incomparably better to be occasionally wrong than always vague. Once granted that inference from experience is ever possible, it is admitted that pure deduction is not the whole of legitimate reasoning, and that the excluded middle is at most a complication to be treated in terms of probability and not a final objection. The procedure is then to state the additional postulates needed as economically as possible. It turns out that traditional mathematics can be extended in scope so as to be applicable to the problems of induction, just as it has been extended to take account of the data that led to the quantum theory. The main result is not any great change from current statistical procedure, but rather that many postulates, introduced in the latter procedure as common-sense statements, but nevertheless apparently independent, can in fact be replaced by consequences of a very few primitive postulates and are therefore closely related.

SUMMARY

Tests are provided for the significance of an estimated departure from a uniform distribution of chance, and of the coefficients of new functions introduced into an empirical law designed to represent a series of measures, in each case where several degrees of freedom may be expected to arise together if one of them does. A test is also given for the independence of errors of observation when the means of groups of consecutive observations

are compared with the standard deviation of the entire set. Applications are made to the secular perturbations of the inner planets and to Pearson's data for errors of observation.

References

Fisher, R. A. 1924 *Proc. Int. Math. Congress, Toronto*, **2**, 805–13.
1928 *Proc. Roy. Soc.* A, **121**, 654–73
— 1935 "The Design of Experiments." Edinburgh: Oliver and Boyd.
1936 "Statistical Methods for Research Workers." Edinburgh: Oliver & Boyd
Jeffreys, H 1935 *Proc Camb. Phil. Soc* **31**, 203–22.
— 1936 *Proc Camb Phil Soc.* **32**, 416–45
— 1937 *a* *Proc Camb Phil Soc* **33**, 35–40
— 1937 *b* "Scientific Inference" Camb. Univ Press
— 1937 *c* *Proc Roy. Soc* A, **160**, 325–48.
— 1937 *d* *Proc Roy Soc* A, **162**, 479–95.
— 1938 *a* *Ann Eugen*, *Camb.*, **8**, 146–51
1938 *b* *Proc. Roy Soc* A, **164**, 307–15
Jeffreys, H. and Bullen, K. E 1935 *Bur Centr Séism. Trav Sci.* **11**, 1–96.
Jones, H. S. 1929 *Observatory*, pp. 297 304.
Pearson, K. 1900 *Phil Mag* **50**, 157–75.
— 1902 *Philos. Trans* A, **198**, 235–99.
Yates, F. 1934 *J. Roy Statist Soc* Suppl 1, pp 217–35.
Yule, G. U and Kendall, M. G. 1937 "Introduction to the Theory of Statistics" London: Griffin.

A new basis for cosmology

By P. A. M. Dirac, F.R.S.
St John's College, Cambridge

(*Received* 29 *December* 1937)

1. Introduction

The modern study of cosmology is dominated by Hubble's observations of a shift to the red in the spectra of the spiral nebulae—the farthest parts of the universe—indicating that they are receding from us with velocities proportional to their distances from us. These observations show us, in the first place, that all the matter in a particular part of space has the same velocity (to a certain degree of accuracy) and suggest a model of the universe in which there is a *natural velocity* for the matter at any point, varying continuously from one point to a neighbouring point. Referred to a four-dimensional space-time picture, this natural velocity provides us with a *preferred time-axis* at each point, namely, the time-axis with respect to which the matter in the neighbourhood of the point is at rest. By measuring along this preferred time-axis we get an absolute measure of time, called the *epoch*.

Such ideas of a preferred time-axis and absolute time depart very much from the principles of both special and general relativity and lead one to expect that relativity will play only a subsidiary role in the subject of cosmology. This first point of view, which differs markedly from that of the early workers in this field, has been much emphasized recently by Milne.

We now feel the need for some new assumptions on which to build up a theory of cosmology. This need is partially satisfied by the assumptions, which Milne calls the Cosmological Principle, that, apart from local irregularities, the universe is everywhere uniform and has spherical symmetry (in three dimensions) about every point, for an observer moving with the natural velocity at that point. The assumption of uniformity is to be taken in its most general form, in which it requires that an observer on another nebula would see all general natural phenomena (for example, the red-shift of other nebulae) the same as we do. The observational evidence in favour of these assumptions is rather meagre, since only a small part of the universe is accessible to present-day telescopes, and this part shows quite large fluctuations from uniformity in the distribution of the spiral nebulae (Reynolds 1937). However, these assumptions are fairly plausible and have a great simplifying effect on the subject, and until there is more definite evidence of their inadequacy it does not seem worth while to try more complicated schemes.

Further assumptions are needed if we are to obtain definite answers to the main problems that suggest themselves in a study of cosmology A possible further assumption is Milne's Dimensional Hypothesis (Walker 1936, p 121), which requires that there shall be no constants with dimensions appearing in cosmological theory. This assumption is open to criticism, as there is no definite reason why the constants of atomic theory should not appear in cosmology---in fact, one would rather expect them to, since one would expect a closer connexion between the atom and the cosmos to show itself with a deeper understanding of Nature. An alternative assumption, which is free from this criticism and is more far-reaching, will be given in the next section and forms the main theme of the present paper.

2. THE FUNDAMENTAL PRINCIPLE

The recession of the spiral nebulae with velocities proportional to their distances from us requires, if we assume these velocities to be roughly constant, that at a certain time in the distant past all the matter in the universe was confined within a very small volume This time appears as a natural origin of time and provides us with a zero from which to measure the epoch of any event. Referred to this zero the present epoch, according to Hubble's data, is about 2×10^9 years.

Let us express this in terms of a unit of time fixed by the constants of atomic theory, say the unit e^2/mc^3. We then get the value 7×10^{38} This turns out to be of the same order of magnitude as the ratio, γ say, of the electric to the gravitational force between an electron and a proton, namely, $2{\cdot}3 \times 10^{39}$. If we had used another atomic unit of time in which to express the present epoch, we should have obtained a value differing from the above one by at most a few powers of ten, which would not have affected the agreement with γ as to order of magnitude, when such large numbers as 10^{39} are concerned. The unit we chose, namely, e^2/mc^3, lies roughly in the geometric mean of all the units of time that we can construct simply from the atomic constants, namely (introducing also the proton mass M),

$$\frac{e^2}{mc^3}, \quad \frac{e^2}{Mc^3}, \quad \frac{h}{mc^2}, \quad \frac{h}{Mc^2}, \quad \frac{\hbar}{mc^2}, \quad \frac{\hbar}{Mc^2},$$

which are in the ratio

1, 0·0005, 850, 0·46, 137, 0·074.

We might have compared the epoch with the ratio of the electric to the gravitational force between two electrons, or between two protons, instead of between one proton and one electron, which would have given us a number 1800 times larger or smaller than γ respectively In any case, however, we see there is a close agreement between the present epoch, expressed in atomic units, and the ratio of the gravitational to the electric force between two elementary particles. Such a coincidence we may presume is due to some deep connexion in Nature between cosmology and atomic theory. Thus we may expect it to hold not only at the present epoch, but for all time, so that, for example, in the distant future when the epoch is 10^{50}, we may expect γ will then be of the order 10^{50}. We are thus led to the result that a quantity γ, usually considered as a universal constant, must vary with the passage of great intervals of time.

A further study of cosmology leads to the appearance of other very large dimensionless numbers. These numbers all turn out to be of the order 10^{39} or sometimes 10^{78}. From a natural extension of the foregoing ideas we should expect all those numbers of the order 10^{39} to increase proportionally to the epoch, and all those of the order 10^{78} to increase proportionally to the square of the epoch We have here a new principle appearing, that all the very large dimensionless numbers occurring in Nature are simple powers of the epoch, with coefficients of the order unity.

To get this principle in its most general form we should not make the assumption, which we made at the beginning of this section, that the velocity of recession of each spiral nebula is roughly constant. Without this assumption we can still talk about the epoch of an event, but we have no natural zero from which to measure it, so that only the difference of two epochs can enter into laws of nature. We must now use Hubble's constant, namely, the coefficient of proportionality between the red-shift and the distance, as one of the quantities from which very large dimensionless numbers are to be constructed (to replace our previous use of the present epoch as one of these numbers) and express our principle in the form· *Any two of the very large dimensionless numbers occurring in Nature are connected by a simple mathematical relation, in which the coefficients are of the order of magnitude unity* If we can deduce from elementary considerations that some of these very large numbers vary with the epoch (as we shall find in the next section is the case), then they must all do so to preserve the mathematical relations between them.

This very general formulation of the principle does not enable one to draw exact conclusions with certainty. If, for example, we have two numbers a and b both of the order 10^{39}, we cannot with certainty conclude that

$$a = kb, \tag{1}$$

where k is a constant of order unity. Owing to our numerical coincidences being inaccurate by a few powers of ten (on account mainly of the uncertainty of which atomic units to use), we must allow k to differ from unity by a few powers of ten, and thus we may have instead of (1), for example,

$$a = kb \log b,$$

with a somewhat different k. In the present paper, for the sake of getting a definite theory, we shall ignore the possible occurrence of such logarithmic factors, or other similar factors that vary slowly with their arguments. We must then remember that the resulting theory will be valid only as a first approximation and may need amendment in the future by the insertion into the equations of functions that vary slowly with their arguments.

Essentially the same approximation is involved in the assumption, which is implied throughout this paper, that hc/e^2 and M/m are constants. Future developments may require these quantities to vary slowly with the epoch.

3. THE LAW OF RECESSION OF THE SPIRAL NEBULAE

Let us take two neighbouring spiral nebulae and express the distance between them in terms of a unit of distance provided by the atomic constants, say the unit of time that we used in the preceding section multiplied by the velocity of light. The distance between the nebulae then becomes a dimensionless number, which will vary with the epoch in an unknown way, and which we call $f(t)$. On account of our assumptions of uniformity and spherical symmetry in § 1, $f(t)$ must be the same for any two neighbouring nebulae, except for an arbitrary constant factor. The determination of the form of $f(t)$, giving the law for the rate of recession of the spiral nebulae, is one of the main problems of cosmology.

Let us obtain Hubble's constant, the red-shift per unit distance, in terms of $f(t)$ The time taken by light to travel from one of our neighbouring spiral nebulae to the other is, since we are using units which make the velocity of light unity, just $f(t)$. If we consider two waves of light starting out from one of the nebulae at times δt apart, they will arrive at the other nebula at times $\delta t + f(t+\delta t) - f(t)$ apart, owing to the different times of transit for the two waves. Thus light which is emitted with the period δt will arrive with the period $\delta t + f(t+\delta t) - f(t)$, and the red-shift, namely, the change in period per unit period, is $\delta f(t)/\delta t = f'(t)$, so that Hubble's constant is $f'(t)/f(t)$. From Hubble's data this has the value at the present epoch $1{\cdot}4 \times 10^{-39}$.

We now bring into the argument the average density of matter ρ, which has a meaning from our assumption of uniformity. We take as unit of mass the mass of the proton or neutron, and we assume that matter is conserved when expressed in this unit. From this assumption of conservation we can infer that, owing to neighbouring nebulae separating from one another according to the law $f(t)$, the average density of matter will decrease according to the law

$$\rho \propto f(t)^{-3} \tag{2}$$

The observed value for the average density of luminous matter is about 5×10^{-31} g. cm.$^{-3}$, which becomes, in our present units, about 7×10^{-45}. This value must be increased by a factor, which is very hard to estimate but is probably a few powers of ten, to get the total average density of all matter. Allowing for the inaccuracy caused by the uncertainty of which atomic units we ought to use, we see that the average density matter is of the same order of smallness as Hubble's constant The reciprocals of these two quantities are two very large numbers, to which our fundamental principle is applicable, and which must therefore be connected like the a and b in equation (1). Thus

$$\rho = kf'(t)/f(t),$$

where k is a constant of the order of magnitude unity. Combining this with (2), we get

$$f(t)^{-3} \propto f'(t)/f(t),$$

and hence

$$f(t) \propto (t+a)^{\frac{1}{3}},$$

a being a constant of integration. By suitably choosing the zero from which we measure t, we may make this constant vanish and we then have

$$f(t) \propto t^{\frac{1}{3}}. \tag{3}$$

This gives us the law for the rate of recession of the spiral nebulae. The velocities of recession are not constant, as we provisionally assumed at the beginning of § 2, but vary proportionally to $f'(t)$ or $t^{-\frac{2}{3}}$. However, with this law of recession we still have a natural origin of time, namely, the zero of the t in (3), when all the nebulae were extremely close together. From (3) we have

$$t = \tfrac{1}{3}f(t)/f'(t), \tag{4}$$

showing that the present epoch is still of the order 10^{39}, and is, in fact, just a third of the value we gave it at the beginning of § 2 on the assumption of constant velocities of recession. The new value, equal to about 7×10^8 years, is rather small, being less than the age of the earth as usually calculated from data of radioactive decay, but this does not cause an inconsistency, since a thorough application of our present ideas would require us to have the rate of radioactive decay varying with the epoch and greater in the distant past than it is now

Our deduction of (3) involves the assumption of conservation of mass when expressed in proton or neutron units, which means conservation of the number of protons and neutrons (apart from processes involving the transformation of the rest-energy of these particles to or from another form). There is no experimental justification for this assumption, since a spontaneous creation or annihilation of protons and neutrons sufficiently large to alter appreciably the law (3) would still be much too small to be detected in the laboratory However, such a spontaneous creation or annihilation of matter is so difficult to fit in with our present theoretical ideas in physics as not to be worth considering, unless a definite need for it should appear, which has not happened so far, since we can build up a quite consistent theory of cosmology without it.

4. THE CURVATURE OF SPACE

Take all the points in space-time for which the epoch has some given value t. They will lie on a three-dimensional surface, which is everywhere orthogonal, in the sense of special relativity, to the natural time-axis. We call it the t-space Our assumptions of uniformity and spherical symmetry in § 1 require that the t-space shall be everywhere uniform and spherically symmetrical. It follows that the t-space must be a *space of constant curvature*, the metric being provided by the atomic unit of distance that we had previously. (For a detailed study of this question, based on group theory, see Walker 1936.)

The curvature may be either positive, zero or negative. If it is positive, the t-space is finite and is like the three-dimensional surface of a sphere in four dimensions. If it is zero or negative, the t-space is infinite and is flat or hyperbolic respectively. Which of these three cases holds cannot, from considerations of continuity, depend on the value of t and must therefore be characteristic of space-time as a whole. To decide between these three cases forms another main problem of cosmology.

The case of positive curvature can easily be ruled out In this case the total mass of the universe is finite and, expressed in the proton or neutron unit, is a very large number. From our assumption of conservation of mass, this large number must be independent of the epoch. We thus get a contradiction with our fundamental principle, according to which all very large numbers occurring in Nature must vary with the epoch, since some of them, namely, the reciprocals of Hubble's constant and of the average density, do.

The case of negative curvature can be ruled out in a similar but rather more complicated way. The total mass of the universe is not finite in this case, but we can work instead with the mass contained at time t within a sphere of radius equal to the radius of curvature of t-space If we take a different epoch t_1, there will be a natural correspondence between points on the t_1-space and points on the original t-space (corresponding points being on the same nebula) and any element of distance in the t_1-space will equal the corresponding element of distance in the t-space multiplied by $f(t_1)/f(t)$. This factor being the same for all the elements of distance, it follows that the radius of curvature of the t_1-space must equal that of the t-space multiplied by this factor. The total mass contained within a sphere of radius equal to the radius of curvature must now be the same for the t_1-space as for the t-space. This mass, expressed in the proton or neutron unit, will again give us a constant number, which must be very large, in order that the curvature of t-space may be sufficiently small not to be in disagreement with observation, and which therefore contradicts our fundamental principle.

We are thus left with the case of zero curvature, or flat t-space, as the only one consistent with our fundamental principle and with conservation of mass. It should be remembered that the curvature we are here speaking about is the curvature of the three-dimensional space at one epoch and not the curvature of space-time as comes into general relativity.

5. THE MOTION OF A FREE PARTICLE

One other problem we shall concern ourselves with in this paper is the determination of the world-line of a particle that is moving freely under the action only of the gravitational field of the universe as a whole. We need something to replace Newton's first law of motion. If the particle is started off with the natural velocity of the place where it is situated, then, from our assumption of the spherical symmetry of the universe about any point, the particle cannot have an acceleration in any direction and Newton's law must hold for it. If, however, it is started off with a different velocity, then we cannot assert more than that its acceleration must lie in the plane in space-time containing its velocity vector and the natural velocity vector of the place. The magnitude of the acceleration may be any function of the velocity of the particle relative to the natural velocity and of the epoch.

Since general relativity explains so well local gravitational phenomena, we should expect it to have some applicability to the universe as a whole. We cannot, however, expect it to apply with respect to the metric provided by the atomic constants, since with this metric the "gravitational constant" is not constant but varies with the epoch. We have, in fact, from the discussion at the beginning of § 2, the ratio of the gravitational force to the electric force between electron and proton varying in inverse proportion to the epoch, and since, with our atomic units of time, distance and mass, the electric force between electron and proton at a constant distance apart is constant, the gravitational force between them must be inversely proportional to the epoch Thus the gravitational constant will be inversely proportional to the epoch.

Let us try to set up a new system of units, whose ratios to the old units may vary with the epoch, so that with respect to the new units the gravitational constant does not vary with the epoch and general relativity may be expected to apply. We must not take a new unit of mass whose ratio to the old one varies with the epoch, as we should then have the mass of a proton or neutron varying with the epoch, and general relativity requires that the mass of an isolated particle shall remain constant. We must therefore change our units of distance and time, and must change both in the same ratio in order to keep the velocity of light unity Since the dimensions of the gravitational constant are $(\text{distance})^3\,(\text{time})^{-2}\,(\text{mass})^{-1}$, we must take new units of distance and time equal to the old ones divided by the epoch, so that the new measure of a distance or time interval is equal to the old one multiplied by the epoch, to make the gravitational constant independent of the epoch.

We may now reasonably assume that, with the metric provided by the new measures of distance and time, general relativity holds and free particles move along geodesics. We thus have two measures of distance and time that are of importance, one for atomic phenomena and the other for ordinary mechanical phenomena included under general relativity. This situation is the same as Milne has with his two measures of time t and τ (Milne 1936, 1937*a*, 1937*b*), but the ratio of the two measures is just the inverse in our theory from what it is in Milne's.

6. THE CURVATURE OF SPACE-TIME

It is of interest to discuss the curvature of space-time, referred to the metric to which general relativity applies, and to determine to what stress-energy tensor it corresponds. This curvature is different, of course, from the curvature of t-space dealt with in § 4 Let us use stars to denote quantities measured in the new units, so that for an interval of time

$$\delta t^* = t\,\delta t$$

and thus

$$t^* = \tfrac{1}{2}t^2, \tag{5}$$

giving us the connexion between the new and the original measure of the epoch. The distance between neighbouring spiral nebulae, expressed in the new units, will vary with the epoch according to the law

$$f^*(t) = t f(t) \propto t^{\frac{4}{3}},$$

and hence

$$f^*(t^*) \propto t^{*\frac{2}{3}}. \tag{6}$$

We may now use Robertson's formulae (Robertson 1933, equations 3·2), according to which the curvature of our space-time must correspond to a uniform density ρ^* and a uniform hydrostatic pressure p^* given by

$$\left.\begin{aligned} \kappa\rho^* &= 3f^{*\prime 2}/f^{*2}, \\ \kappa p^* &= -2f^{*\prime\prime}/f^* - f^{*\prime 2}/f^{*2}, \end{aligned}\right\} \tag{7}$$

where κ is the constant of gravitation and the primes indicate differentiations with respect to t^*. We are here taking Robertson's k equal to zero, since our t-space is flat, and we are taking the cosmical constant λ occurring in Einstein's law of gravitation to be zero, since if it were not zero it would have to be very small not to be in disagreement with observation and its reciprocal would then provide us with a very large constant number, in contradiction to our fundamental principle.

Substituting (6) into (7), we get

$$\kappa\rho^* = \tfrac{4}{3}t^{*-2} = \tfrac{16}{3}t^{-4}, \tag{8}$$

$$\kappa p^* = 0. \tag{9}$$

From (2) and (3) we have

$$\rho \propto t^{-1},$$

which is in agreement with (8) when one remembers the different units of distance used in the measurement of ρ and ρ^*. This agreement should not be regarded as a support for our present theory, however, since it is due simply to the observed average density of matter being of the same order of magnitude as that to be expected from the curvature of space-time (assuming a radius of curvature of the order of the reciprocal of Hubble's constant), which fact provides a satisfactory feature in every theory of cosmology. On the other hand the result (9) may be regarded as a support for our theory, since the average hydrostatic pressure in space, due mainly to radiation pressure, is extremely small compared with the average density of matter, and so should be counted as zero in a first approximation.

SUMMARY

It is proposed that all the very large dimensionless numbers which can be constructed from the important natural constants of cosmology and atomic theory are connected by simple mathematical relations involving coefficients of the order of magnitude unity. The main consequences of this assumption are investigated and it is found that a satisfactory theory of cosmology can be built up from it

REFERENCES

Milne, E A 1936 *Proc. Roy. Soc.* A, **158**, 324–48.
— 1937*a* *Proc. Roy. Soc.* A, **159**, 171–91.
— 1937*b* *Proc. Roy. Soc.* A, **159**, 526–47
Reynolds, J H 1937 *Nature, Lond*, **140**, 387.
Robertson, H P 1933 *Rev Mod Phys* **5**, 62–90.
Walker, A. G. 1936 *Proc Lond Math Soc.* **42**, 90–127

The scattering of cosmic ray particles in metal plates

BY P. M. S. BLACKETT, F.R.S. AND J. G. WILSON

(*Received* 27 *January* 1938)

1. THE EXPERIMENTAL RESULTS

Using an ordinary cloud chamber, Anderson (1933) measured the average angle of scattering of cosmic-ray particles of energy up to 3×10^8 e-volts in a lead plate of thickness 1·1 cm Williams (1936) pointed out that, in such experiments, the scattering is multiple, and that the observed values agreed with the theoretical predictions, assuming the particles to be electrons. Recently Neddermeyer and Anderson (1937) have made some new measurements using a counter-controlled cloud chamber and a 1 cm. platinum plate. No numerical results are given, but the observed scattering of particles of energy up to 5×10^8 e-volts, which appears to be the limit of the energy measurements, seems in rough agreement with the earlier results.

Using the counter-controlled cloud chamber already described by Blackett (1936) and by Blackett and Brode (1936), the multiple scattering of cosmic-ray particles of energy up 9×10^9 e-volts in lead and copper plates has been measured. In order to make such measurements possible, it is necessary to reduce as far as possible the distortion of the tracks in the chamber. The technique by which this can be achieved has been described by Blackett and Wilson (1937) in connexion with the measurement of the energy loss of rays in traversing metal plates. In fact, the photographs taken for the measurement of the energy loss were, amongst others, found suitable for the measurement of the scattering. The angle of scattering was measured by means of a goniometer eyepiece attached to a travelling microscope. The cross-wires were set tangentially to the track, first on one side and then on the other side of the plate. The mean of the two measured angles of deflexion, obtained from the two stereoscopic photographs, was taken as giving the projection on the plane of the chamber of the actual angle of scattering.

This measured deflexion θ_m of each track was corrected for three sources of error. The first is the random error of the angle measurements, due to random chamber distortions and setting errors An upper limit of 0·16° of arc for the probable error due to this cause was obtained by assuming that the rays of high energy ($H\rho > 3 \times 10^7$ gauss cm.) suffered no real scattering, so that all the observed scattering was taken as due to errors.

The second is a small systematic deflexion due to chamber distortion, which amounted to about 0·04° of arc. Both these are small compared with the measured mean angles of scattering, except for the highest energies.

The third error is the systematic one due to the curvature C of the track in the field.* If the two angle settings be made at a distance d apart, the change of direction of the track between the two points due to the field is $\theta_1 = Cd$. This error is additive or subtractive, according as the sign of the measured angle θ_m of scattering is in the opposite, or in the same, direction as the deflexion θ_1 due to the curvature. For a 1 cm plate, d is about 2 cm., but its exact value depends somewhat on the particular observer's method of setting the cross-wires. For any group of tracks of neighbouring energy, the value of θ_1 can be obtained directly from the measurements of the mean apparent scattering angle. It is simply given by the algebraic mean of the measured deflexions θ_m of all the tracks, these being taken as, say, positive, when θ_m and θ_1 are of the same sign, and negative when they are of opposite sign Since the observed mean scattering is found to be nearly proportional to the curvature of the track (that is inversely proportional to the energy), it follows that the above correction is nearly the same fraction of the mean scattering angle at all energies. Though this fraction is about 40 % for the tracks traversing 1 cm. Pb in 10,000 gauss, the uncertainty in the mean scattering angle due to the uncertainty in the magnitude of the correction θ_1 is quite small, certainly less than 5 %, since in about half of any group of tracks the correction is additive and in the other half must be subtracted.

The results are given in Table I. The first column gives the range of energy of the group of tracks, calculated from $E = 300\ H\rho$ e-volts, that is, assuming their mass to be electronic. The value of $H\rho$ for each track is taken as the mean of the values above and below the chamber. The second column gives the harmonic mean of the energies of the tracks in the group, followed by the corresponding mean values of $H\rho$, and then by the number of tracks. The fifth column gives the arithmetic mean angle Θ of the measured deflexions, corrected for the three sources of error. The last two columns give the theoretical value α, and the ratio of the observed to the theoretical value respectively.

* This source of error can be avoided, in principle, by the method of Simons and Züber (1937). However, this method is only easily applicable in practice to the tracks of low energy.

TABLE I. SCATTERING OF COSMIC RAYS IN METAL PLATES

Range of E 10^9 e-volts	$\bar{E}$ 10^9 e-volts	$\overline{H\rho}$ 10^7 gauss-cm	No.	Θ degrees of arc	α degrees of arc	Θ/α
1 cm. lead, 10,000 gauss:						
0·15–0·45	0·29	0·097	7	2·96 ± 0·65	2·80	1·06 ± 0·24
0·45–0·90	0·59	0·197	15	1·40 ± 0·22	1·38	1·02 ± 0·16
0·90–1·35	1·09	0·36	13	0·96 ± 0·15	0·75	1·28 ± 0·20
1·35–2·1	1·71	0·54	19	0·55 ± 0·14	0·48	1·15 ± 0·30
2·1 –3·3	2·54	0·85	15	0·56 ± 0·09	0·32	1·75 ± 0·28
3·3 –9·0	4·65	1·55	15	0·31 ± 0·06	0·18	1·72 ± 0·35
> 9·0	—	-	21	0·16	—	—
0·33 cm. lead, 3300 gauss:						
0·03–0·10	0·07	0·023	8	7·45 ± 1·9	6·55	1·14 ± 0·29
0·1 –0·4	0·21	0·070	9	1·77 ± 0·60	2·12	0·84 ± 0·29
0·4 –0·8	0·55	0·18	9	0·62 ± 0·10	0·82	0·76 ± 0·12
2 cm. copper, 10,000 gauss:						
0·1 –1·0	0·32	0·11	15	2·18 ± 0·31	2·05	1·08 ± 0·15
1 2	1·25	0·42	10	0·56 ± 0·13	0·53	1·05 ± 0·25
2 –4	2·72	0·91	14	0·48 ± 0·09	0·24	2·00 ± 0·37

2. COMPARISON WITH THEORY

Williams (1936, 1938) has shown that the theoretical value of the arithmetic mean projected angle α of scattering of a particle of mass M, and unit charge, and velocity v very nearly c, in a plate of thickness t, which contains N atoms, of atomic number Z, per cm ^{3}, is given approximately by

$$\alpha^2 = \frac{4Z^2e^4Nt}{\xi^2M^2C^4}\log(2\pi NtZ^{4/3}\lambda_0^2), \tag{1}$$

where $\xi = (1 - v^2/c^2)^{-\frac{1}{2}}$, and λ_0 is the Compton wave-length. Since

$$H\rho = Mc^2\xi/e, \text{ when } \xi \gg 1,$$

we have

$$\alpha = \frac{Ze(Nt)^{\frac{1}{2}}}{H\rho}\log(2\pi NtZ^{4/3}\lambda_0^2), \tag{2}$$

and Williams points out that this expression does not depend on the mass of the particle. It is seen that the mean angle of scattering should be inversely proportional to the measured $H\rho$ of the rays.

The values of α calculated from (2) are given in Table I. In the last column is given the ratio Θ/α of the observed to the calculated angle of scattering. Since these values are all of the order of unity, it can be concluded that the theoretical expression (2) is roughly verified for both lead and copper. For lead, this rough verification extends over a range of $\overline{H\rho}$ of 60 to 1. The corresponding range of energies, assuming the mass of the particles to be electronic, is from 7×10^7 to $4{\cdot}6 \times 10^9$ e-volts.

It can be seen from fig. 1, which shows the results graphically, that the agreement between theory and experiment is good for values of $\overline{H\rho}$ up to about 6×10^6 gauss-cm., but that for higher energies the observed values are greater than those calculated. At these higher energies the energy measurements are least accurate, and so the possibility of some unexpected source of error is greatest. One might therefore reasonably look for the discrepancy in the systematic errors in the measured energies due to the random errors of measurement. But the discussion of § 6·2 of the paper by Blackett and Wilson (1937) suggests this effect is too small, and probably also of the wrong sign.

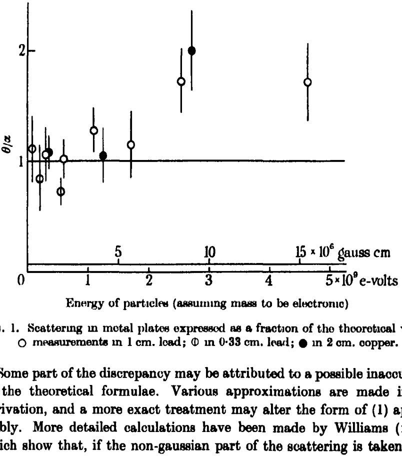


FIG. 1. Scattering in metal plates expressed as a fraction of the theoretical value. ○ measurements in 1 cm. lead; ⦶ in 0·33 cm. lead; ● in 2 cm. copper.

Some part of the discrepancy may be attributed to a possible inaccuracy of the theoretical formulae. Various approximations are made in its derivation, and a more exact treatment may alter the form of (1) appreciably. More detailed calculations have been made by Williams (1938) which show that, if the non-gaussian part of the scattering is taken into account, and if the shielding of the atomic electrons is allowed for by the use of a Fermi-Thomas atomic field, the theoretical values are increased by about 20 %. This correction seems too large at the lower energies and not large enough at the higher energies, in order to bring agreement between theory and experiment.

It is not thus possible to say with any certainty whether the apparent excess of the observed scattering of the rays of high energy over that predicted is real or not. Further experiments with a 2 cm. gold plate are in progress to test this. If it proves to be real it is possible that it is associated with the process of energy loss shown by Wilson (1938) to be of particular importance for energies of the order of 10^9 e-volts.

Fig. 2 shows the distribution of the measured angles of scattering in a 1 cm. lead plate of all rays with $H\rho < 3 \times 10^6$ gauss-cm. In view of (2), it is the distribution of the product $\theta \times H\rho$, and not that of θ alone, which is plotted. The distribution is approximately Gaussian as would be expected.

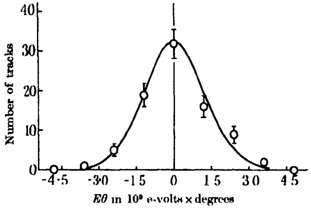


FIG. 2. Distribution of corrected deflexions of all tracks with $H\rho < 3 \times 10^6$ gauss-cm. The abscissae are not the corrected deflexions, but the products of the deflexion by the energy of the particles.

3. DISCUSSION

Williams (1938) has pointed out that if the high penetration of the majority of the cosmic rays is explained by the assumption made by Neddermeyer and Anderson (1937) that the particles are heavier than electrons, one would still expect the normal scattering given by (2). Thus the fact that nearly the theoretical scattering is observed gives support for the view that it is by their greater rest mass that the penetrating rays are distinguished from radiating electrons. Williams further gives an argument based on the method of impact parameters, which leads to the view that any other way of reducing the radiation loss will also reduce the scattering and so bring disagreement with experiment. If this argument is valid it becomes certain that the mass of the penetrating rays must be greater than that of electrons.

It is interesting to note that the scattering of the rays shows no noticeable change at an energy of about 2×10^8 e-volts, although it has been shown by Blackett and Wilson (1937), and by Blackett (1938), that the mean energy loss of the rays changes very rapidly at about this energy. The particles whose scattering has been measured in the present work are almost all of the penetrating group, since the number of energetic electrons at sea-level is very small (Blackett 1938)

Since Blackett and Wilson (1937) and Blackett (1938) have shown that the penetrating rays become electronic in character, when their energy falls below about 2×10^8 e-volts, it follows that the explanation of the remarkable change of property in this energy region must be sought, on this view, in a *change* of rest mass. But since it does not seem likely that the present quantum theory is adequate to describe this supposed change of rest mass, it will probably prove necessary to introduce some new physical principle to explain it, and this new principle may possibly invalidate the argument, which is based on existing theory, that a normal scattering and an abnormal radiation loss can only be explained by means of a heavy particle. Until therefore an adequate explanation of the change of radiative property has been found, it would be perhaps premature to conclude with certainty that the difference between the penetrating and the absorbable rays can only be explained by a difference in rest mass, though this explanation seems at present to be the most plausible.

We wish to express our appreciation of the work of Mr A. H. Chapman, who took all the photographs. This work was carried out at Birkbeck College, London.

SUMMARY

1. Measurements have been made of the multiple scattering of cosmic rays in the following metal plates: 0·33 cm lead, 1·0 cm. lead, 2·0 cm. copper. The range of values of $H\rho$ of the tracks extended from 10^5 to 3×10^7 gauss-cm., corresponding to electron energies of 3×10^7 to 9×10^9 e-volts.

2. The observed average angle Θ of the multiple scattering is found to be nearly inversely proportional to the measured values of $H\rho$, and to be

in fairly close agreement with the prediction of theory for rays of any mass, but with velocity nearly equal to c.

At high energies the observed values are somewhat higher than expected. This discrepancy, which may be partly due to experimental error, will be further investigated.

3. This result, that the scattering of the penetrating component is normal, while the radiation loss is much less than that expected for electrons, gives support for a heavier rest mass for the rays. But this conclusion is not quite certain

REFERENCES

Anderson 1933 *Phys Rev.* **43**, 381
Blackett 1936 *Proc Roy. Soc.* A, **146**, 83
— 1938 *Proc. Roy. Soc.* A, **165**, 11.
Blackett and Brode 1936 *Proc. Roy Soc* A, **154**, 573.
Blackett and Wilson 1937 *Proc Roy. Soc* A, **160**, 304.
Neddermeyer and Anderson 1937 *Phys. Rev* **51**, 884.
Simons and Zuber 1937 *Proc. Roy Soc.* A, **159**, 383
Williams 1936 "Kern Physik", p. 123. Berlin· Springer.
— 1938 *Phys Rev.* (in the Press).
Wilson 1938 *Proc Roy. Soc.* A (in the Press).

On the motion of a fluid heated from below

By R. J. Schmidt and O. A. Saunders
Imperial College of Science and Technology

(*Communicated by S. Chapman, F R S.—Received* 11 *November* 1937)

[Plates 1, 2]

Introduction

In some previous experiments (Schmidt and Milverton 1935) a layer of water between two horizontal plates was slowly heated from below The critical temperature difference at which the water began to move was found from a change in the slope of the curve relating the difference of temperature between the plates and the rate of supply of heat to the lower plate. An optical refraction (Saunders and Fishenden 1935) method was also used for finding the critical condition, and the results found by the two methods agreed and conformed to a theoretical formula of Jeffreys (1928) within the limits of experimental error.

The present experiments were undertaken to find whether any change in the type of motion occurs at higher temperature differences, and also to study further the vertical and horizontal temperature gradients in the moving fluid using the optical method. It was also thought of interest to perform experiments with air instead of water. An improved apparatus was used, in which the downward heat loss from the lower plate could be measured, so that the actual heat transfer between the plates could be found.

Description of apparatus

The apparatus consisted essentially of two square brass plates supported horizontally one above the other. The plates were supported separately by glass rods, glass being used as it has approximately the same thermal conductivity as water. The bottom plate was heated electrically whilst the top plate was cooled by water flow. The temperatures of the opposing surfaces of the plates were measured by embedded thermocouples.

The bottom plate A (see fig. 1) measured $9 \times 9 \times \frac{3}{8}$ in., and had a thermocouple (not shown in figure) embedded in a deep groove cut along its under surface, the junction being situated just below its upper surface and approximately at the centre of the square face. The plate was held by five fibre screws to an ebonite slab *B*, $\frac{3}{8}$ in. thick, with a heating coil of resistance tape sandwiched between. Thin sheets of mica were used to insulate the heating coil on either side, and these were allowed to project about $\frac{3}{4}$ in. all round the edge of the plate. By dipping the projecting edges in molten paraffin wax and afterwards scraping away the excess wax flush with the upper surface, the plate was thermally insulated at the edges and at the same time made watertight. Pairs of embedded thermocouples gave the temperatures of the two surfaces of the ebonite at three positions (not shown in the figure) from which the mean rate of conduction of heat through the ebonite could be estimated with sufficient accuracy. The whole was supported on three glass rods fixed to a wood base. All electrical leads passed through plugs on the underside of the ebonite and through watertight rubber tubing.

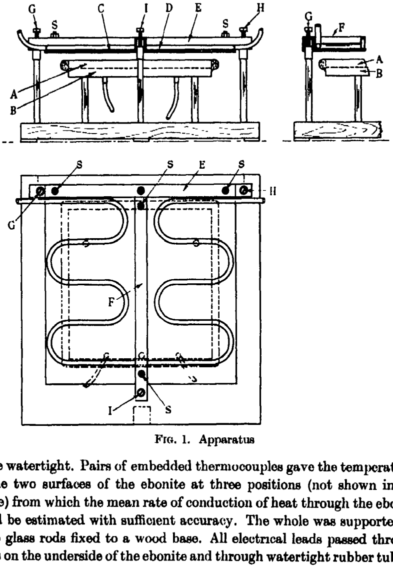


FIG. 1. Apparatus

The top plate *C* measured $12 \times 12 \times \frac{1}{4}$ in. and had an embedded thermocouple similar to that in the bottom plate at its centre, and a copper cooling coil D soldered to its upper surface. Brass bars, E and F, of inverted U section, were also attached by screws S to its upper surface, and projected a short distance beyond the edges of the plate, at the positions G, H and I. The plate was supported by three vertical screws passing downwards through threaded holes in the bars at G, H and I, with their tips resting on three brass caps fitted over glass rods fixed to the wood base. The height and level of the plate could thus be adjusted by the screws.

Both plates had flat surfaces, finished by scraping and testing on a surface plate, and their distance apart when in contact varied from point to point by less than 0·01 cm.

The whole apparatus stood in a tank with vertical glass sides. In the experiments with water the tank was filled to a level just above the under surface of the top plate.

EXPERIMENTAL PROCEDURE

In an experiment, after levelling the bottom plate by adjusting three screw-jacks supporting the tank, three conical distance pieces of height equal to the required distance between the plates were placed on the bottom plate The top plate was then lowered, by means of its supporting screws, until it just rested on the distance pieces. After noting the positions of the adjusting screw heads, the top plate was raised by turning them slightly, so that the distance pieces could be removed, and the plate was returned to its previous position At the beginning of an experiment the water in the cooling coil attached to the top plate was at the same temperature as the fluid in the tank. The heating current was switched on and, after waiting 45 min., which was found to be long enough for the flow of heat to become steady, temperature readings were taken. The current was then given its next value and the proceedings repeated. The results of four experiments are given in Table I, in which d is the distance between the plates, C is the measured current, θ_2 and θ_1 are the temperatures of the surfaces of the top and bottom plates respectively, and θ_{1e} and θ_{2e} are the mean temperatures of the upper and lower surfaces of the ebonite. The resistance of the heating coil was 38·0 ohms, and was constant within $\frac{1}{4}$ % over the range of temperatures used; thus the total heat supplied in unit time to the lower plate was proportional to C^2.

DISCUSSION OF RESULTS

The results from Table I are plotted in fig. 2, the co-ordinates being the square of the heating current and the temperature difference between the

TABLE I

C (amp)	θ_1° C.	θ_2° C.	θ_{1e}° C.	θ_{2e}° C.
		$d = 1·1$ cm.		
0·40	19·05	18·4	19·8	18·7
0·50	19·6	18·7	20·6	18·6
0·605	20·1	18·9	21·4	18·5
0·725	20·6	19·0	22·5	18·6
0·81	21·05	19·1	23·4	18·6
0·91	21·65	19·35	24·6	18·7
1·03	22·4	19·65	26·0	18·9
1·175	23·5	20·3	28·1	19·2
1·415	25·4	21·15	31·85	19·8
1·71	28·0	22·4	37·5	20·6
		$d = 1·0$ cm.		
0·40	17·9	17·2	18·8	17·7
0·55	18·65	17·75	20·1	17·6
0·71	19·95	18·5	22·0	17·9
0·82	20·65	18·85	23·3	17·9
0·95	21·45	19·1	25·0	18·1
1·11	22·55	19·6	27·1	18·4
1·28	23·9	20·2	29·7	18·8
1·47	25·6	21·15	32·8	19·3
1·70	27·7	22·2	37·3	20·1
		$d = 0·9$ cm.		
0·40	17·0	16·55	17·5	15·8
0·55	18·0	17·2	19·0	16·0
0·77	19·5	18·1	21·7	16·4
0·97	21·0	18·7	24·5	17·0
1·21	23·05	19·65	28·2	17·6
1·41	24·6	20·3	31·4	18·3
1·66	26·85	21·4	35·8	19·1
1·92	29·3	22·65	41·0	19·9
2·17	32·1	23·9	46·5	21·0
		$d = 0·8$ cm.		
0·40	16·85	16·2	17·8	16·7
0·61	18·1	17·0	20·1	16·9
0·82	19·6	17·8	22·9	17·2
1·02	21·15	18·5	25·7	17·7
1·17	22·4	19·05	28·2	18·0
1·39	24·2	19·95	31·4	18·5
1·57	25·9	20·7	34·7	19·0
1·88	29·05	22·4	40·5	19·9
2·235	33·3	24·65	48·7	21·2

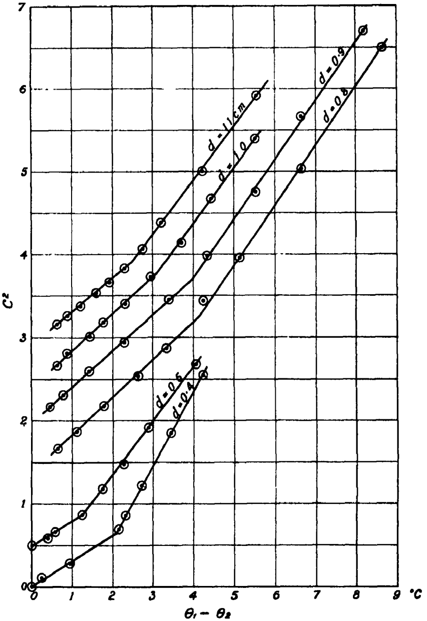


FIG. 2. Experimental results for water, showing critical temperature differences.

Note. The zero of the C^2 scale has been displaced upwards by respectively 0·5, 1·5, 2·0, 2·5 and 3·0 units for $d = 0{\cdot}5$, 0·8, 0 9, 1·0 and 1·1 cm.

plates. It will be seen that each curve has a change in slope at a critical temperature difference, which decreases when the distance apart is increased. No allowance has been made in fig. 2 for the heat which passes downwards through the ebonite, which amounted to about one-sixth of the total heat supplied to the bottom plate, for the purpose of detecting changes in the slope of the curve relating the heat transfer and the temperature difference between the plates, it was found unnecessary to correct for the ebonite loss. Table II shows the values of $\lambda \equiv \alpha g c(\theta_1-\theta_2)d^3/k\nu$ (Jeffreys 1928) corresponding to each of the critical temperature differences. In this expression g is the acceleration due to gravity, α, c, ν and k are respectively the expansion coefficient, the specific heat per unit volume, the kinematic viscosity and the thermal conductivity of the water, taken at the mean of the temperatures of the two plates.

TABLE II

d	1·1	1·0	0·9	0·8	0·5	0·4
θ_1	21·75	22·7	24·0	24·15	18·35	19·8
θ_2	19·35	19·6	20·0	19·95	17·1	17·7
λ	46,000	47,000	47,000	35,000	1800	1700

The results for $d = 0{\cdot}8$, $0{\cdot}9$, $1{\cdot}0$ and $1{\cdot}1$ cm. appear to indicate a change in the type of motion at a critical value of λ equal to about 45,000. The use of the optical method, which will be described later, confirms this, and shows that the steady cellular motion, which occurs at lower values of λ, breaks down at about $\lambda = 45{,}000$ and turbulent motion begins, in which the water temperatures fluctuate with time

The first appearance of cellular motion, which was the subject of investigation in a previous paper (Schmidt and Milverton 1935), occurs at temperatures below the smallest experimental values in Table I. For $d = 0{\cdot}4$ and $0{\cdot}5$ cm., however, the change of slope is clearly shown in fig 2, the corresponding critical values of λ being 1800 and 1700 respectively (see Table II). This agrees with the results in a previous paper It is to be noted that the change of slope at 1700 is sharper than that at $\lambda = 45{,}000$. It was thought unnecessary to give detailed results for $d = 0{\cdot}4$ and $0{\cdot}5$ cm. in Table I, since these experiments are repetitions of those already given in the paper referred to.

Results for heat-transfer with water

The mean rate of heat loss through the ebonite beneath the bottom plate was calculated for each experiment, assuming a value 0·00042 cal./cm. °C. sec. for the conductivity of the ebonite, and subtracted from the total energy supplied. Thus H, the rate of heat transfer per unit time per unit surface area of the bottom plate was found, neglecting edge losses from it. The results obtained are given in fig. 3, in which the dimensionless co-ordinate $\lambda Hd/k(\theta_1-\theta_2)$ is shown plotted against λ. It was thought desirable to plot the product of λ and $Hd/k(\theta_1-\theta_2)$ against λ, rather than $Hd/k(\theta_1-\theta_2)$ against λ, since the product does not involve $(\theta_1-\theta_2)$ and gives a better grouping of the experimental results, because the accuracy in measurement of $\theta_1-\theta_2$ is less than that of H and d.

The results for different values of d fall reasonably well on one curve, although there is a slight spread at higher values. The change of slope at $\lambda = 45{,}000$ is, however, not so well shown up as when the points for each value of d are plotted separately.

OPTICAL METHOD

Light from a carbon arc, placed 13 m. away from the centre of the fluid layer, in a direction parallel to one of the sides of the plates and in the same horizontal plane as the layer, passed through the fluid and was intercepted by a screen some distance away on the opposite side. Negative photographs of the image so formed were taken by placing specially sensitive negative cards over the screen and exposing for about 2 sec

Since the refractive index of the fluid decreases with increase of temperature, the light is bent upwards in passing through the layer. Let x, y, z, denote distances measured parallel to the horizontal sides of the square plates bounding the layer and perpendicular to them, respectively. All rays from the source, which pass through the layer, may be regarded as entering in the direction of the x-axis. If the angular deviation is everywhere small, the small angle made with the xy plane by a ray on leaving the layer is approximately proportional to the mean z-gradient of refractive index taken along its path. This angle may be found from the vertical displacement of the image from its position when the temperature of the layer is uniform. Similarly, the displacement of the image parallel to y is approximately proportional to the mean y-gradient of refractive index taken along the path of the ray Further, assuming the refractive index to vary linearly with temperature, as was approximately true in the present experiments, these displacements measure the corresponding mean gradients of temperature.

A number of negative photographs are shown in figs. 4 and 5, the appropriate values of d, $\theta_1-\theta_2$, λ, and l, the distance between the arc and the apparatus, being given. The shaded lines at the edges of the photographs

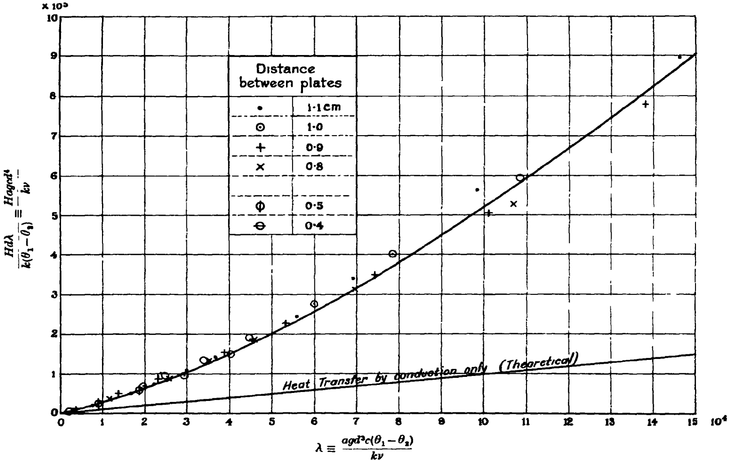


FIG 3 Heat transfer across water between plates.

show the positions of the images formed by the light when the temperature is uniform.

Results for water by the optical method

The negative photographs for water are shown in fig. 4, Plate 1.

Photograph (*a*). By means of a screen containing three horizontal slits placed between the arc and the apparatus, light was admitted only within small distances of the top surface, the mid-plane, and the bottom surface, of the layer, respectively. When the given temperature difference was applied, the photograph shows that the light passing through the middle slit is practically undeflected vertically, except near the edges, while the light from both the top and the bottom slits is deflected upwards, the deflexion varying periodically with distance across the width of the photograph. The maxima of vertical deflexion of the light from the bottom slit occur immediately above the minima of vertical deflexion of the light from the top slit; these positions indicate where the cooled liquid is descending. The image formed by the light from the top slit lacks some of the rays which entered the layer close to the top surface, which are either absorbed or reflected by the surface; this accounts for the average deflexion of the light from the top slit being apparently less than that of the light from the bottom slit.

The absence of vertical deflexion of all rays passing near the mid-plane shows that the vertical temperature gradients are negligible everywhere on this plane. The transfer of heat across this plane must therefore be practically all by convection.

Photographs (*b*) *and* (*c*). When obtaining these photographs light was admitted to the whole gap between the plates. The image when the layer was at a uniform temperature thus consisted of a wide band, whose position is shown at the side. The characteristic features are the bright vertical patches, which are more marked in (*b*) than in (*c*) owing to the greater distance between the layer and the screen and the correspondingly sharper focussing. They occur at the same positions as the maxima of the vertical deflexion of the light passing close to the bottom surface, which forms the second wavy line from the top. This is to be expected since the light is deflected horizontally towards the positions where the cooler fluid is descending. There are no vertical patches in photograph (*a*) because the temperature gradients are smaller than in (*b*) or (*c*), despite the larger value of λ.

In (*a*) there are eleven cells, and in (*b*) and (*c*) twenty-two cells, within a horizontal length of 22·9 cm.; hence the horizontal lengths of the sides of the cells are 2·1 and 1·0 cm., respectively, i.e. approximately twice the distance between the plates. It was generally found that when the cellular

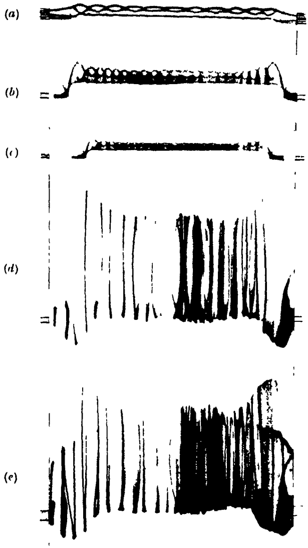


FIG. 4 Photographs with water.

(*a*) $d = 1{\cdot}1$ cm., $\lambda = 12{,}000$, $\theta_1 - \theta_2 = 0{\cdot}55^\circ$ C., $l = 500$ cm.
(*b*) $d = 0{\cdot}5$ cm., $\lambda = 3500$, $\theta_1 - \theta_2 = 1{\cdot}7^\circ$ C., $l = 500$ cm.
(*c*) $d = 0{\cdot}5$ cm., $\lambda = 3500$, $\theta_1 - \theta_2 = 1{\cdot}7^\circ$ C., $l = 250$ cm.
(*d*) $d = 1{\cdot}1$ cm., $\lambda = 130{,}000$, $\theta_1 - \theta_2 = 4{\cdot}0^\circ$ C., $l = 500$ cm.
(*e*) $d = 1{\cdot}1$ cm., $\lambda = 130{,}000$, $\theta_1 - \theta_2 = 4{\cdot}0^\circ$ C., $l = 500$ cm.

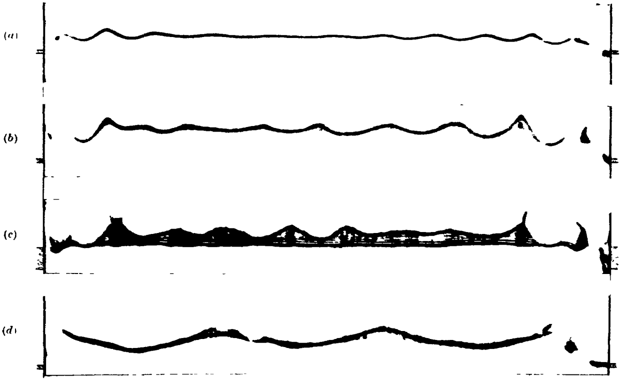


FIG. 5 Photographs with air

(*a*) $d = 1 \cdot 0$ cm, $\lambda = 3400$, $\theta_1 - \theta_2 = 34°$ C.
(*b*) $d = 1 \cdot 0$ cm, $\lambda = 7000$, $\theta_1 - \theta_2 = 86°$ C.
(*c*) $d = 1 \cdot 0$ cm, $\lambda = 6500$, $\theta_1 - \theta_2 = 74°$ C.
(*d*) $d = 2 \cdot 0$ cm, $\lambda = 96{,}000$, $\theta_1 - \theta_2 = 101°$ C.

motion was well established the width of a cell was about $2d$. According to the theory, if the cells are square their sides are of length $2{\cdot}8d$ when the motion first begins. Theoretically, however, it is possible for smaller cells to form as the motion develops, although according to Rayleigh (1916) the cells first formed have the best chance of predominating. Alternatively the discrepancy would be explained if square cells formed with their sides parallel to the diagonals of the plates.

In (*b*) and (*c*), unlike (*a*), the bottom parts of the images, which correspond to rays passing near the mid-plane, show considerable vertical deflexion, proving that there are considerable vertical temperature gradients across the mid-plane. Since the value of λ in (*b*) and (*c*) is lower than that in (*a*), it appears that as the motion develops with increasing λ, the vertical temperature gradients across the mid-plane decrease, and therefore less conduction takes place across the mid-plane. This is well shown, although it was not appreciated at the time, in fig. 4, photographs (*c*), (*d*), (*e*) of the previous paper to which reference has already been made. These photographs were obtained with circular plates, and the vertical deflexion thus varies from a maximum at the middle of the photographs to zero at the edges on account of the varying length of path within the layer; but it will be seen that the bottom of the image in (*e*) is straighter than that in (*d*) which is in turn straighter than that in (*c*); the corresponding values of λ increased in the order (*c*), (*d*), (*e*).

Photographs (*d*) *and* (*e*). These photographs were taken when the fluid was in turbulent movement, with an exposure of 1 sec. In the left half of each photograph the light entering the layer was limited to eleven vertical slits 0·2 cm. wide and 1·0 cm. apart. In the right half of each photograph no vertical slits were used and the vertical lines in the image are due to refraction in the fluid. Further, in (*d*) the light entering the layer was restricted to the lower half of the gap between the plates, the light entering the upper half being cut off by a screen. In each case the image formed when the fluid temperature was uniform thus consisted of a broad horizontal band, whose position is shown at the right of each photograph, the band being interrupted by the vertical slits in the left half of each photograph and appearing as a series of short vertical lines. Disregarding the turbulent change between the two exposures, (*e*) thus consists of (*d*) together with the image formed by the light entering the upper half of the gap, some of which is reflected from the top plate and can be seen faintly at the bottom of (*e*). As before the bottom parts of the images are formed by light entering near the mid-plane, and these show that the vertical temperature gradients are practically negligible on this plane except near the edges. From the left halves of the photographs it is seen that the horizontal displacements, which are proportional to the horizontal temperature gradients, are small at the top and the bottom, which correspond to rays passing respectively near the plates and near the mid-plane, but are larger at intermediate positions. When the image was observed on a screen it was seen that the tops and bottoms of the vertical lines in the left half of each photograph remained practically stationary while the intermediate parts continually moved from side to side. A typical bow-like shape of line is seen just to the left of the middle of each photograph. The periodic time of the fluctuations could not be said to be at all definite, but was of the order of one second.

The appearance of vertical lines in the right halves of (*d*) and (*e*) is due to focusing of light by horizontal deflexion towards descending columns of cooled water. These lines continually shifted about. Their general bow-like shape can be explained in a similar manner to the lines in the left halves of the photographs.

Results for air by optical method

The rate of change of refractive index with temperature is smaller for air than for water, but by increasing the distance between the apparatus and the screen, it was found that the optical method could be used satisfactorily. The experiments were carried out with the same apparatus and experimental arrangements as for water, but it was found necessary to enclose the gap between the plates by vertical glass walls in order to get any image at all, presumably because air in cellular motion is more sensitive than water to disturbances outside the layer.

Fig. 5, Plate 2 shows the image obtained for four different conditions. In every case the apparatus was 13 m. from the arc, and the image 17 m. from the apparatus. The temperature of the top plate was about 21° C., and the exposure times were all about 1·5 sec.

In (*a*), (*b*) and (*d*) the light entering the layer was restricted to within 0·3 cm. of the bottom plate. In (*c*) the light was admitted over the whole gap between the plates. The images formed when the air was at a uniform temperature thus consisted of straight horizontal bands, at positions indicated at the sides of the photographs.

Photograph (*a*). Cellular motion is clearly indicated, there being eleven cells, the same number as for water with the same distance apart of the plates.

Photograph (*b*). Cellular motion is again indicated, but the image was observed to change slightly with time. The number of cells remained the same, but whereas in the photograph the wavy line is better defined in the right-hand side of the picture, after half a minute or so the left-hand side would probably show better definition. It was not possible to observe a definite time period in the movement of the image.

Photograph (*c*). This was taken for conditions similar to those of (*b*), and shows similar features to fig. 4 for water.

Photograph (*d*). When this was taken the motion had become turbulent and the image was moving about all the time, it will be seen to have moved during the 1·5 sec. exposure. The vertical deflexion showed a general tendency to vary periodically across the horizontal width of the photograph but the "wave-length" was rather greater than for the earlier stages of the cellular motion, in which there would be five or six complete cells for this value of *d*.

A large number of photographs similar to figs. 4 and 5 were taken for different conditions and the following conclusions were drawn.

For low values of λ the appearance of the image with air was similar to that with water at the same value of λ. With air, as with water, the image first showed cellular motion when λ was just under 2000.

With air the image showed first signs of unsteadiness at about $\lambda = 5000$, the movement becoming more marked as λ increased; the number of cells remained the same but they showed a tendency to shift in position. With water, on the other hand, the image was fairly steady until λ equalled about 45,000 when turbulence set in more or less suddenly. With air the turbulent movement developed more gradually, and at $\lambda = 45{,}000$ the amount of turbulence was very roughly the same as with water shortly after turbulence first appeared.

The first appearance of cellular motion in air when λ equals about 2000 is, of course, in agreement with the theory.

The experimental result that turbulence develops at about the same value of λ (45,000) for both air and water is of interest. Although this condition has not been studied mathematically, an examination of the equations of motion of the fluid, which must involve the inertia terms, indicates that the critical conditions should depend upon both λ and $k/c\nu$.

Since the values of $k/c\nu$ appropriate to the experiments were approximately 1·3 for air and 0·14 for water, the above result appears to indicate that the amount of turbulence depends mainly on λ.

Attempts to detect changes of motion of the air by plotting the heat supplied to the lower plate against the temperature difference between the plates were less satisfactory than for water, the change of slope being more difficult to locate.

The authors wish to thank Professor Newall for his interest in the work. They are also indebted to the Clothworkers Company for financial assistance towards the cost of apparatus.

SUMMARY

If a horizontal layer of fluid, initially at rest, is heated from below, instability and cellular motion set in when λ ($\equiv \alpha gc(\theta_1-\theta_2)d^3/k\nu$) is about 1700 Experiments for several layer depths show that on further heating the layer becomes completely turbulent at about $\lambda = 45{,}000$. Observations by an optical method indicate that in the case of water the turbulence develops suddenly, but that in air the transition is more gradual, the first signs of turbulence occurring at about $\lambda = 5000$. The rate of heat transfer is measured at each stage, and the increase at the change from cellular to turbulent motion is found to be less than that at the change from equilibrium to cellular motion.

In the cellular motion the length of the horizontal side of a cell is found to be twice the layer depth. The vertical temperature gradients at positions on the plane midway between the boundary surfaces decrease progressively to negligible values as the motion develops with increasing λ

In the turbulent motion the fluctuating vertical temperature gradients are very small on the mid-plane. The horizontal gradients are very small on the mid-plane, but have maxima at positions between the mid-plane and either boundary surface.

REFERENCES

Jeffreys, H. 1928 *Proc Roy Soc.* A, **118**, **195**.
Rayleigh, Lord 1916 *Collected Papers*, **6**, 438
Saunders and Fishenden 1935 *Engineering*, **139**, **483**.
Schmidt and Milverton 1935 *Proc. Roy. Soc.* A, **152**, **586**.

Identification and measurement of helium formed in beryllium by γ-rays*

BY E. GLÜCKAUF AND F. A. PANETH

(*Communicated by J. C. Philip, F R S —Received* 17 *December* 1937)

For reasons explained elsewhere (Paneth 1937) helium is at present the only product of artificial disintegration which can be separated in bulk by chemical methods. Some time ago helium artificially produced by bombardment of boron with slow neutrons was spectroscopically identified and measured by a manometric device (Paneth, Glückauf and Loleit 1936). This reaction

$$ {}^{10}_{5}\mathrm{B} + {}^{1}_{0}n \rightarrow {}^{7}_{3}\mathrm{Li} + {}^{4}_{2}\mathrm{He} \qquad (1)$$

had been investigated before by physical methods (Chadwick and Goldhaber 1935*a*, *b*; Taylor and Goldhaber 1935; Amaldi and others 1935). The fact of the production of helium was already known and chemistry could merely contribute to its quantitative study. There are, however, other processes of artificial disintegration where at present only microchemical methods seem capable of deciding between different possibilities.

Since the discovery of the "nuclear photo-effect" by Chadwick and Goldhaber, it is known that γ-rays produce neutrons from deuterium and beryllium (Chadwick and Goldhaber 1934, 1935*c*; Szilard and Chalmers 1934), in this case the formation either of ^{8}Be or of ^{4}He is to be expected according to the equation

$$ {}^{9}_{4}\mathrm{Be} + h\nu \rightarrow {}^{8}_{4}\mathrm{Be} + {}^{1}_{0}n \qquad (2)$$

or

$$ {}^{9}_{4}\mathrm{Be} + h\nu \rightarrow 2\,{}^{4}_{2}\mathrm{He} + {}^{1}_{0}n \qquad (3)$$

Each reaction is equally possible, although Chadwick and Goldhaber give reasons why reaction (2) is probably the main one. Now if reaction (3) occurs to any considerable extent, then it should be possible to detect the helium by gas-analytical methods, while in the case of formation of a stable, or slowly disintegrating, ^{8}Be, no helium should be found.

Owing to the minute quantity of helium to be expected even after irradiation with strong γ-ray sources, it is essential to absorb as much as possible of the γ-radiation in beryllium atoms; on the other hand, the larger the amount of irradiated material, the more difficult becomes the chemical separation of the helium. For these reasons the ideal arrangement is a radioactive source, as small as possible, and surrounded by a sphere of metallic beryllium. It is true that the collection of the helium from the metal is more difficult than from a soluble beryllium salt; but a simple calculation shows that in order to disintegrate the same number of beryllium atoms in a dissolved beryllium salt, one would have to deal with impracticable quantities of solution. On the other hand, in the case of the irradiation of a dry salt, with subsequent solution, the complete exclusion of air would be difficult. It was decided, therefore, to irradiate metallic beryllium, and to develop a method for the quantitative recovery of the helium formed.

---

* "Helium Researches XIV." References to the previous numbers are given in "Helium Researches XIII" (Paneth, Glückauf and Loleit 1936).

APPARATUS AND EXPERIMENTAL PROCEDURE

Analyses, to be described later, showed that commercial beryllium does not contain any detectable traces of helium; certainly less than 10^{-10} c.c./g. beryllium. It could, therefore, be used for experiments without re-melting in vacuo. As all metals are perfectly helium-tight, it is unnecessary to enclose the beryllium in an evacuated vessel during the irradiation; on the other hand, the only possible way of collecting the helium after the irradiation is to dissolve the metal. The considerable volume of hydrogen evolved in this operation causes the only serious difficulty in the whole experiment. The methods used hitherto for the separation of small quantities of helium from a large amount of hydrogen were either the absorption of hydrogen by calcium, or the diffusion of the gas through heated palladium ("Helium Researches, I", fig. 1, p 358 and fig. 2, p. 360, Paneth and Peters 1928). To be efficient the latter method needs a very large palladium spiral, which was not at our disposal, while the calcium method depends largely on the presence of traces of other metals in the surface of the calcium and seldom works satisfactorily after the absorption of a few litres of hydrogen. Therefore in the present experiment the hydrogen was burned with oxygen, the mixture being ignited by a heated wire. In this way it is possible to remove up to 20 l. of hydrogen in about an hour; the last traces of the gas are best burned in the palladium furnace already described ("Helium Researches, XIII", fig. 1, p. 416, Paneth, Glückauf and Loleit 1936).

Since a reliable method for the separation of minute quantities of helium from many litres of hydrogen is not only essential for our present experiment, but may be useful in similar cases, it seems advisable to give an accurate description of the procedure, with a few words about the preparation of the air-free acid and the dissolution of the beryllium.

The apparatus (fig. 1) consists mainly of a storage vessel for the acid *A*, from which the liquid can be siphoned into the flask *B*, containing the beryllium metal. A lump of beryllium which has not been irradiated and is to be used for the blank test, lies on the bottom of the flask, while the irradiated piece, fixed with copper wire to a small iron rod, is held by means of the magnet *C* in the neck of the flask.

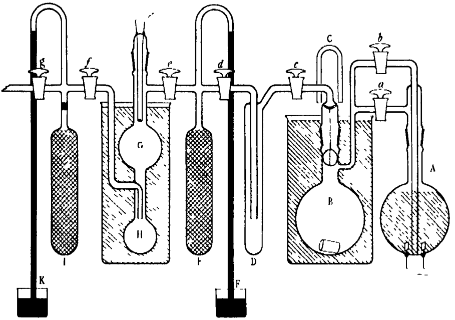


Fig. 1

The storage bulb *A* is filled with 20 % sulphuric acid which, though forming a less soluble beryllium salt, is preferable to nitric acid because it avoids the attack on the mercury of the manometers by nitrous fumes. It is also superior to hydrochloric acid on account of its lower volatility. The sulphuric acid is electrolysed for 2 hr. under about 20 mm. Hg, it is then completely free from air which would otherwise introduce contaminating helium and neon. After this purification the cooled trap *D*, in which a few c.c. of water have been condensed, is warmed up, and the water freed from occluded gases by boiling under reduced pressure. The water is then frozen out again and the charcoal tubes *F* and *I* freshly activated by heating. Sufficient oxygen for the combustion of the hydrogen is electrolytically prepared and purified, as described in the previous publications, and stored in the charcoal tube *I* connected with the manometer.

The stopcock *a* is now closed and the acid siphoned into the flask *B* under the pressure of the electrolytic gas. As the dissolution of the beryllium produces considerable heat, it is necessary to surround the flask *B* with water, occasionally cooled by the addition of ice. The hydrogen evolved is dried in the trap *D* (cooled by liquid air), and stored in the charcoal tube *F* (likewise immersed in liquid air). A manometer *E* allows the hydrogen pressure to be read, and at the same time constitutes a safety valve through which the hydrogen escapes if the pressure exceeds 1 atmosphere.

When the pressure of hydrogen reaches about 300 mm., the stopcock *e* is slightly opened so as to admit it slowly into the combustion chamber *G*, filled with oxygen at 150–200 mm pressure. (It is essential not to exceed this pressure, as otherwise an explosion might occur at the moment of ignition.) When the hydrogen, containing of course some oxygen, passes the heated platinum spiral at the top of the combustion chamber *G*, ignition occurs, and, by regulating the stopcock *e*, it is possible to maintain the flame burning in this chamber, which is cooled outside by water. It is of importance to keep the flame in the upper half of the spherical chamber, as only there is it stable; if the hydrogen pressure decreases, the flame withdraws into the narrower space, which causes the rate of combustion to be retarded on account of the reduced surface available for the contact between hydrogen and oxygen If, on the other hand, the hydrogen pressure increases, the flame advances towards the middle of the bulb until it reaches an area wide enough to cope with the hydrogen supplied through the stopcock *e* It is obvious that the flow of hydrogen must be regulated in such a way that the flame never reaches the equator of the spherical combustion chamber, otherwise it at once becomes unstable and moves with great velocity through the narrow tubes towards the oxygen-filled charcoal tube *I*. (This would result in an explosion, and as a precautionary measure a plug of glass-wool is introduced into the tube leading to *I*.) The flame emits only a very dim reddish light, hence for its control it is necessary to carry out the combustion in a darkened room.

After some hydrogen has been burned away, the liquid air vessels which cool the charcoal tubes *F* and *I* are gradually lowered until practically all the hydrogen is burned to water, which forms in bulb *H*. The water is boiled in order to remove any helium which might possibly have been included during the process of condensation, and the remaining gases, mainly oxygen, are transferred through stopcock *g* into the analysing apparatus. This has already been fully described ("Helium Researches, XIII", fig. 2, p. 417, Paneth, Glückauf and Loleit 1936).

After the blank test with ordinary beryllium has been satisfactorily carried out, the magnet *C* is removed, so that the irradiated piece of beryllium drops into the acid. The whole procedure is then repeated.

The experiment was carried out three times. For irradiation, pieces of beryllium, more or less spherical and weighing 6·6, 3·75 and 4·5 g., were used, and as γ-ray source radon (supplied by the Radium Centre of Middlesex Hospital). The latter was contained in glass capillaries of an average length of 6 mm. and 1 mm. width, the quantity of the radon in each capillary varying between 100 and 250 mc. Generally the Middlesex Hospital supplied a tube once a week, two or three of them could be simultaneously placed inside the beryllium in holes drilled towards the centre of the metal. After the decay of the main part of the radon, the tubes were removed and replaced by fresh ones. As there is always some danger that the radon tubes are contaminated outside with radioactive material or may leak slightly, they were covered with other capillaries before insertion In addition to this precaution the surface of the beryllium holes was removed by drilling, and by washing with acid, before the chemical analysis was undertaken.

RESULTS*

The results of the experiments can be seen from Table I. It comprises two blank tests, carried out with pieces of the same beryllium samples as were used for the irradiation experiments. In none of the blanks could as much as 10^{-10} c.c of helium/g. of beryllium be detected.

TABLE I HELIUM IN METALLIC BERYLLIUM AFTER BOMBARDMENT WITH γ-RAYS

Quantity of Be, in g.	Time of irradiation, in days	Radon decayed, in mc.	He found in 10^{-8} c.c.	He/g. Be, in 10^{-8} c.c.	He/g. Be/ mc. Rn, in 10^{-12} c.c.	He/cm. Be/ mc. Rn in 10^{-11} c.c.
8·1	0	0	<0·1	<0·01	—	—
6·6	26	600	2	0·3	5	3·7
3·75	22	1250	2·6	0·7	5·5	2·8
0·8	0	0	<0·01	<0·01	—	—
4·5	260	4700	18	4·0	8·5	4·8

Each of the three irradiated pieces of beryllium contained an amount of helium which could easily be detected and approximately measured, namely, 2×10^{-8}, $2·6 \times 10^{-8}$ and 18×10^{-8} c c. While in the first two cases, owing to the small quantity of helium, the manometric measurement was only about

---

* A preliminary report was published in *Nature* (Paneth and Glückauf 1937); only two of the three experiments were described therein.

25 % accurate, in the last the measurement itself was reliable to ± 10 %; unfortunately, owing to a slight accident, a fraction of the hydrogen, mixed with helium, had escaped during the dissolution of the beryllium. As this fraction is known only approximately, the final figure in this experiment is only accurate within 20 %. The values "He/g. Be" are given for comparison with the helium content of non-irradiated beryllium, and of beryls (see the following paper). Since the beryllium samples were approximately spherical, it is possible from their weight to calculate roughly the helium content per 1 cm. radius of the beryllium spheres surrounding the radon source These values—which should be independent of the quantities of beryllium used—are given in the last column. The average is $3{\cdot}8 \pm 0{\cdot}8 \times 10^{-11}$ c.c. helium/cm. beryllium/mc. radon decayed.

MEASUREMENT OF THE NUMBER OF PHOTO-NEUTRONS BY THE QUANTITY OF HELIUM PRODUCED IN BORON

The amount of helium in the three pieces of irradiated beryllium which must thus be ascribed to the γ-rays is so considerable that it would appear to be formed by the main reaction; for if one assumes the γ-rays of radium C of 1·8 and 2·2 million V to be equally efficient, and the cross-section of the beryllium nucleus to be 5×10^{-28} cm.[2], a quantity of helium corresponding to the order found is to be expected.* Because of the uncertainties in these assumptions, however, an attempt was made to decide by a more exact method whether helium is the main product of the reaction. Should this be so, the number of helium atoms produced in beryllium would be twice that of the photoneutrons (see reaction (3)); this latter figure can be determined, as shown in our last paper, from the equal number of helium atoms produced in boron according to reaction (1). With this aim in mind the third experiment was arranged in the following way.

The sphere of metallic beryllium of 0·84 cm. radius (weight 4·5 g.) was placed during the whole irradiation in the pocket of the methyl borate vessel of 7·5 cm. radius, described in "Helium Researches, XIII". In the experiments on the formation of helium from boron, for the bombardment neutrons produced by the impact of the α-particles of Rn, Ra A, and Ra C on beryllium were used; of these neutrons, in spite of the slowing down effect of the hydrogen of the methyl borate, a considerable fraction was not caught inside the copper vessel. In the present experiment, however, the initial velocity of the neutrons was much smaller, owing to their character as photo-neutrons; therefore it could reasonably be expected that most of them would be absorbed by the boron inside a vessel of 7·5 cm. radius. It was thought advisable, however, to verify this and at the same time to determine the number of neutrons originating from the glass of the radon tubes under the impact of the α-rays; the amount of helium which these neutrons produce in the boron, must, of course, be subtracted for the purpose of comparison.

---

* According to recent experiments the cross-section is four times as large (Frisch, v. Halban and Koch 1937), but this does not change the order of the helium quantity to be expected.

For these experiments Dr P. B. Moon kindly lent us an ionization chamber, filled with boron fluoride and connected with a counter constructed by Dr C. E Wynn-Williams. With this device it was observed that the number of neutrons produced in a radon-filled glass capillary and slowed down by paraffin amounted to 20 % of the neutrons obtained when the glass capillary was inserted in the centre of our beryllium metal. If such a complex source is placed inside the water-filled pocket of the boron vessel and surrounded by a large water tank, the absorption of those neutrons which owe their origin to the impact of α-particles on glass must amount to roughly 45 %; for this is the absorption measured under the same conditions for neutrons produced by the impact of α-rays on beryllium. Of the total number of neutrons 82 % were caught in the boron vessel, since 20 % of these neutrons are 45 % absorbable, the absorption of the photo-neutrons alone consequently amounts to about 91 %

As, therefore, 9 % of the photo-neutrons escape, this number must be added to the helium atoms found in the boron; but as, on the other hand, 45 % of the 20 % glass neutrons are also used for helium production in boron, the same negative correction has to be applied. It follows from this consideration that the *number of helium atoms found in the boron* can (within the accuracy attainable in our experiments) be considered as *equal to the number of neutrons produced by the γ-rays in the beryllium.*

The analysis of the helium content of the boron vessel was carried out in the apparatus described in "Helium Researches, XIII"; the only improvement was the attachment of a reflux condenser to the boron vessel. With an arrangement as shown in fig. 2 of "Helium Researches, XIII", the boiling of the boron ester results in the distillation of 5–10 % into the trap; its passage through the metal stopcock attacks the grease. In the new experiment, between the vessel and the stopcock a metal reflux condenser (see fig. 2) was inserted. As cooling agent during the analysis trichloroethylene was used, cooled by solid carbon dioxide contained in the funnel; the trichloroethylene, heated by the boiling ester, rises in the condenser and is, in constant flow, cooled down by the carbon dioxide.

The analysis of the helium content of the methyl borate vessel, after the beryllium sphere in the pocket had been irradiated by γ-rays for 8·5 months, showed the presence of $11 \pm 3 \times 10^{-8}$ c.c. helium. In the beryllium (see above) $18 \pm 4 \cdot 10^{-8}$ c.c. helium had been found. Now we know that the number of helium atoms found in the boron is practically equal to the number of photo-neutrons produced in the beryllium. The number of helium atoms found in the beryllium is more than 1·6 times as high; if reaction (3) were the only one induced by the γ-rays in the beryllium, the ratio of the helium atoms to the neutrons in the beryllium should be 2. The result of our experiments therefore makes it clear that the amount of helium produced in the beryllium by γ-irradiation stands in such proportion to the number of photo-neutrons as is to be expected if ^{4}He, and not a stable ^{8}Be, is the main final product of the reaction.

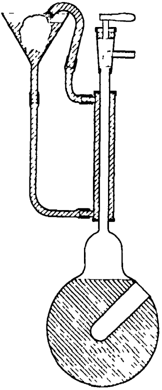


FIG. 2.

A repetition of the experiment with a greater quantity of radon, a larger vessel containing methyl borate, and a few additional refinements would make it possible to decide with an accuracy of 5–10 % whether there is any formation of a stable ^{8}Be. Even such an increased accuracy will not, however, permit the exclusion of the possible intermediate formation of an unstable ^{8}Be. Since the irradiation of the beryllium, with the radon sources at our disposal, has to be continued for several months before sufficient helium is produced, a ^{8}Be with a half-value period of no more than 1 month would have time practically completely to decay according to the reaction

$$^8_4\text{Be} \rightarrow 2\,^4_2\text{He}. \tag{4}$$

Such a disintegration of the ^{8}Be into two helium atoms may take place spontaneously without detectable rays, since within present-day accuracy the mass of ^{8}Be is twice the mass of ^{4}He For several reasons—Chadwick and Goldhaber's indirect conclusions, and Bohr's general picture of the mechanism of the nuclear disintegration—it seems more likely that the photo-disintegration of ^{9}Be proceeds according to reaction (2) and not to reaction (3); if one accepts this view, it follows from our experiments that ^{8}Be is not stable but splits up into 2 ^{4}He. Dr Goldhaber has kindly informed us that according to unpublished experiments of Mrs Fremlin and Dr Kempton—in which ^{8}Be recoil atoms from the reaction $^9\text{Be} + {}^1\text{H} \rightarrow {}^8\text{Be} + {}^2\text{D}$ were collected inside a Geiger counter—the lifetime of ^{8}Be is shorter than 1 sec. or longer than 3 years. Combined with the experiments here described we may therefore conclude that the lifetime of ^{8}Be is shorter than 1 sec.

We gratefully acknowledge our indebtedness to the Imperial College of Science and Technology for laboratory facilities, to Imperial Chemical Industries, Ltd., for financial help, to Professor S. Russ, director of the Radon Department, Middlesex Hospital, London, for supplying the radon tubes; and to Dr M. Goldhaber, Cavendish Laboratory, Cambridge, for suggesting the experiment and for helpful discussions.

SUMMARY

While it has been known for some time that the beryllium nucleus, irradiated by γ-rays, emits neutrons, it could not be decided whether the nucleus is thereby transformed into a stable isotope of beryllium of mass 8, or into two helium atoms.

By a micro-chemical method helium was detected in beryllium after irradiation by the γ-rays of radon In one of the experiments the number of neutrons simultaneously emitted during the irradiation was determined by measuring the helium produced by these neutrons in methyl borate; a comparison of the two helium quantities showed that the main final product of the γ-irradiation of beryllium is helium and not the beryllium isotope.

REFERENCES

Amaldi, D'Agostino, Fermi, Pontecorvo, Rasetti and Segrè 1935 *Proc. Roy. Soc.* A, **149**, 522
Chadwick and Goldhaber 1934 *Nature, Lond.*, **134**, 237.
— 1935 *a* *Nature, Lond.*, **135**, 65.
— 1935 *b* *Proc Camb. Phil. Soc* **31**, 612.
— 1935 *c* *Proc. Roy. Soc.* A. **151**, 479.
Frisch, v. Halban and Koch 1937 Copenhagen Conference. (Not yet published.)
Paneth 1937 *J. Chem. Soc.* p 642.
Paneth and Gluckauf 1937 *Nature, Lond.*, **139**, 712
Paneth, Gluckauf and Loleit 1936 *Proc. Roy. Soc.* A, **157**, 412.
Paneth and Peters 1928 *Z. phys. Chem.* **134**, 353.
Szilard and Chalmers 1934 *Nature, Lond.*, **134**, 494.
Taylor and Goldhaber 1935 *Nature, Lond.*, **135**, 341.

On the occurrence of helium in beryls*

BY J. W. J. FAY, E. GLÜCKAUF AND F A. PANETH

(*Communicated by J. C. Philip, F.R.S.—Received* 17 *December* 1937)

When Lord Rayleigh first discovered the occurrence, in samples of beryl, of helium greatly in excess of the amount attributable to the traces of uranium and thorium contained therein, he suggested as the most plausible hypothesis that an unknown element present in beryl may emit α-particles with less than the critical velocity (Strutt 1908). Later, on the suggestion of B. B. Boltwood, he discussed another possibility, viz. that in crystallizing from the rock magma the beryls had occluded one of the shorter lived α-radioactive elements such as radium or ionium, which later decayed, leaving nothing but the helium as evidence of its former presence (Strutt 1910).

In view of the subsequent discovery of isotopy both hypotheses were reconsidered in the light of this new knowledge. It has, for example, been suggested that the "unknown element", decaying without detectable rays, might well be the isotope ^{8}Be; and further, that this isotope might have already completely decayed (Atkinson and Houtermans 1929; Rayleigh 1929).

---

* "Helium Researches, XV." For previous papers of the series see preceding contribution (Glückauf and Paneth 1938) and "Helium Researches, XIII" (Paneth, Gluckauf and Loleit 1936; a footnote in the latter gives references to numbers I–XII).

Further, following the discovery of artificial transmutation, it seemed possible to attribute the origin of helium in the mineral to the influence of external rays to which it had been exposed during geological times; this latter idea has been supported in particular by O. Hahn, who thinks that the discovery by Szilard and Chalmers (1934, Chadwick and Goldhaber 1935) of the nuclear photo-effect in beryllium provides a completely satisfactory explanation for the occurrence of helium in beryls, because, after the expulsion of a neutron from ^{9}Be, the remaining ^{8}Be might split up into two ^{4}He (Hahn 1934, 1935, 1936, Born 1936).

A number of experiments have already been carried out in order to decide whether disintegration of beryllium, spontaneous or induced, can be the cause of the helium production. Langer and Raitt (1933) thought they had discovered a permanent α-radiation of the element beryllium, but their findings could not be substantiated by Evans and Henderson (1933), Lord Rayleigh (1933 *a*), Gans, Harkins and Newson (1933), Libby (1933), Fränz and Steudel (1934), and Burkser *et al.* (1937). Chadwick and Goldhaber (1935), and Bernardini and Mando (1935) have shown that even under the bombardment of γ-rays beryllium does not emit an observable α-radiation.

All these experiments suffer from the disadvantage that they cannot detect α-rays emitted with an energy under a value of about 100,000 V. Since it is known that, within the accuracy of present-day measurements, ^{8}Be has twice the mass of ^{4}He, a disintegration according to the reaction

$$^8\text{Be} \rightarrow 2\,^4\text{He} \tag{1}$$

would be practically rayless. Therefore the absence of detectable α-rays is in no way decisive. A direct test for helium, on the other hand, is entirely independent of the energy of reaction (1). For this reason it could be hoped, with the help of our sensitive micro-method, definitely to determine whether disintegration of beryllium can explain the helium content of beryls.

I

Let us first consider whether the γ-rays from minerals and rocks in the neighbourhood of beryls can be held responsible for the occurrence of helium, as suggested by O. Hahn.

The experiments described in "Helium Researches, XIV" show that a γ-irradiation of beryllium actually leads to the formation of helium and not of stable "beryllium eight"; so qualitatively the hypothesis is upheld. But, at the same time, they demonstrate that the effect is not nearly sufficient to account for the quantity of helium found, because Lord Rayleigh (1933 *b*) showed that in Archaean beryls the helium content may amount to 77·6 cu. mm./g. Since it is not impossible that helium has escaped in course of time from the mineral, this figure is a minimum. As 1 g. of beryl contains at the utmost 5 % of the element beryllium, it means that 1·55 c.c. of helium/g. Be has been formed since the solidification of the mineral, which must certainly have occurred less than 2000 million years ago. The average annual production of helium would, therefore, have to be at least $7·7 \times 10^{-10}$ c.c./g. of Be.

It has been found (see preceding paper) that 1 mc. radon which decays in the centre of a sphere of beryllium of a few grams weight produces therein about $6·5 \times 10^{-12}$ c.c. of helium/g. of beryllium. The decay of 1 mc. radon is equivalent to the constant γ-radiation of 1 mg. of radium during 5·52 days; in 1 year 1 mg. of radium would, therefore, produce $4·3 \times 10^{-10}$ c.c. For the formation of $7·7 \times 10^{-10}$ c.c annually, about 1·8 mg. of radium would require to be present (If the same calculation is made for the beryl from Latah Co. which contains 2·27 cu. mm. of helium/g., and belongs to the Mesozoic (Rayleigh 1933 *b*)—age < 200 million years—we find that 0·53 mg. radium would have to be present.) Furthermore, since radium cannot maintain its activity for more than a few thousand years, we should need to assume the presence of its parent substance uranium in the proportion of 3×10^6 g./g. radium; that would mean 5·4 kg. of uranium element or 6·4 kg of U_3O_8/g. of beryllium. If, lastly, it is taken into account that radium, uniformly distributed in an uranium mineral, is necessarily much farther distant from the beryllium in a beryl than the radon capillary from the beryllium metal in our experiments (average distance 4 mm.), it follows that, owing to the double effect of distance and absorption of the rays, the necessary intensity of γ-radiation could only be produced if every gram of beryl were surrounded by many tons of a uranium—or thorium—mineral. These considerations render Hahn's suggestion quite unacceptable.

Further, the proof that no appreciable amount of helium can be formed in beryllium by the γ-radiation of the surroundings, according to present-day knowledge, altogether excludes this radiation as the source of helium in beryls, for in no other elements than beryllium and deuterium with "natural" γ-rays has a nuclear photo-effect been observed (Chadwick and Goldhaber 1935).

It is true that Hahn also considers a possible influence of cosmic radiation, in addition to the γ-radiation from the rocks As this cosmic radiation is constantly present in laboratories, its influence on the helium production would have made itself felt in the experiments on the spontaneous formation of helium in beryllium, to be described later. Furthermore, it is obvious that to assume a marked effect of cosmic radiation on the helium formation in beryls would involve impossible suppositions about the cross-section of the beryllium atoms for the absorption of the radiation.

II

The first experiments intended to detect a spontaneous helium formation in beryllium were begun eleven years ago (Paneth and Peters 1928); after beryllium nitrate solutions had been standing for two years, it was found that air had not been sufficiently rigidly excluded, and these experiments were, for other reasons, never completed. Two years ago experiments were initiated, in which a number of large glass bulbs containing beryllium nitrate solution were set up, freed from air, and allowed to stand for some months, protectively immersed in water. Similar bulbs, some containing ammonium nitrate and others sodium nitrate, were used as blanks. As deducible from previous experiments, helium was found even in the blanks—although this time air leakage was avoided—for glass, being permeable to atmospheric helium, always contains absorbed helium. As the ratio of the release of helium from the large glass bulbs was of the same order as the effect to be investigated, these experiments were abandoned. For such researches either metal vessels must be employed (Paneth, Gluckauf and Loleit 1936) for the solutions to be studied, or else the use of solutions must be completely avoided. The latter method of attack, with pieces of beryllium metal, was employed in the experiments described in the preceding paper.

As therein pointed out, the fact that this metal is helium-tight, obviates the necessity of enclosing it during the course of the supposed helium formation. Any pieces of the metal can, after sufficient lapse of time, be made use of, the older specimens being, of course, the more valuable for the purpose. Unfortunately the metal has only been manufactured on a large scale during the last few years, so that suitable samples of greater age are comparatively rare. Thanks, however, to the courtesy of several industrial companies and of colleagues, some fairly old samples were obtained. The results of the analyses are given below (Table I).

As all the values are maxima, it is obvious that the lowest result is the most important, but in order not to rely on one experiment we shall consider only the average of the four; this shows that 1 g. of beryllium yields less than $1{\cdot}3 \times 10^{-11}$ c.c. of helium per year. It may be emphasized that the samples were of course subject to the influence of cosmic radiation throughout their existence.

This figure is sufficiently low to permit of several inferences. It is at once evident that nowadays at least there is no helium production in beryllium which could account for the quantities found in beryls; for these an average annual production of $7{\cdot}7 \times 10^{-10}$, i.e. 60 times as much, is necessary (p. 240). Hence beryllium exhibits neither α-radiation, nor helium production. To-day it seems to be a stable element.*

TABLE I. HELIUM IN METALLIC BERYLLIUM

Origin of Be	Minimum age, in years	Quantity, in g.	Minimum age × g.	Spectrum	Upper limit of He, in 10^{-10} c.c.	Upper limit of He/g./year, in 10^{-11} c.c.
Comp. de Produits Chimiques, Paris	(1 month old)	8·5	—	Ne + He	1	—
Johnson-Matthey, London	1·5	8·1	12·2	Ne	2	1·7
Fränkel and Landau, Berlin	3	7·1	21·3	—	1	0·5
Comp. de Produits Chimiques, Paris	4	2·0	8·0	Ne + He	3	4
Professor E. Wiberg, Karlsruhe	5	5·7	28·5	Ne + He	3	1
			70		9	Av. 1·3

Just as the hypothesis has been advanced that the "unknown" radio element may have decayed in the meantime, one can argue that while the production of helium from ^{8}Be was appreciable in former geological times it has now ceased (Rayleigh 1929). The following calculation would seem to dispel even this idea; it is based on the helium content of a comparatively young beryl, because then it can be shown that the maximum half-value period is so short that impossibly great quantities of the ^{8}Be would have had to be present at the time of the origin of the earth.

If the helium in the beryl came from ^{8}Be, at the time of formation of the mineral the amount of ^{8}Be must have been at least equal to $\frac{a}{1-e^{-\lambda t}}$, where a is the weight of the helium still present in the mineral, λ the disintegration constant of ^{8}Be, and t the time elapsed since formation of the beryl. On the other hand, to-day 1 g. of beryllium yields per year less than a definite

---

* For the sake of completeness it may be added that beryllium does not spontaneously emit any β-rays either (Friedländer 1935), and that in an experiment with a boron fluoride chamber it was ascertained that 1 g. of beryllium does not emit as much as 1 neutron every 20 min.—a result to be expected from theoretical considerations; for the reaction

$$^{9}\mathrm{Be} + h\nu \rightarrow {}^{8}\mathrm{Be} + n$$

an energy of $1{\cdot}6 \times 10^{6}$ e-volts is necessary (Chadwick and Goldhaber 1935).

quantity of helium (b). From these two data can be calculated the maximum half-value period of ^{8}Be according to the formula

$$b = \lambda \frac{a}{1-e^{-\lambda t}} e^{-\lambda t} = \frac{a\lambda}{e^{\lambda t}-1}.$$

In the beryl from Latah Co. ($t < 2 \cdot 10^{8}$ years) Lord Rayleigh found 2·27 cu. mm. of helium/g., i.e. $8 \cdot 1 \times 10^{-6}$ g./g. ^{9}Be ($=a$). In our experiments it was observed that 1 g. of beryllium yields less than $1 \cdot 3 \times 10^{-11}$ c.c. of helium/year, i.e. $2 \cdot 3 \times 10^{-15}$ g /year ($=b$). From this follows a minimum λ of $2 \cdot 2 \times 10^{-8}$, or a maximum half-value period T of $3 \cdot 2 \times 10^{7}$ years. From a and λ is obtained for the original amount of ^{8}Be, 2×10^{9} years ago, 6×10^{11} g. ^{8}Be for each g. of ^{9}Be, an absurdly high value for an unstable isotope of the comparatively rare element beryllium.*

CONCLUSIONS

It is deduced that the external radiation to which the mineral beryl is subject is not adequate to account for the quantities of helium occluded. Further, the spontaneous decay of beryllium, including the effect of cosmic radiation, produces so little helium, if indeed any, that at no time could this phenomenon have been sufficient to explain the facts.

It seems, therefore, probable that the helium does not come from the beryllium at all but from some other source. A further argument in favour of this is the fact that the amount of helium found in different beryllium-bearing minerals shows no proportionality to the percentage of beryllium present (Strutt 1908; Paneth and Peters 1928); a recent statement (Burkser, Kapustin and Kondogury 1937) to the contrary is based on only one selected sample of beryl. Further, Lord Rayleigh's tables show enormous variations, with only a very rough correlation to time; and although it is possible that unequal amounts of helium have been lost according to the differences in temperature to which the beryls have been exposed, it seems more likely that the variations of the helium content depend upon their original chemical composition. The role of the beryl is perhaps only that of hindering the escape of the helium; it may be that other minerals containing the same helium-producing element are not equally well able to retain the gas.

From the fact that a large helium content is limited to specimens of great geological age Lord Rayleigh has already concluded that it cannot have been generated by the decay of any relatively short-lived radioactive constituent, which may have been initially present. We may add that his figures seem to show that neither is the helium due to the decay of a radio-active element whose half-value period is long compared with the age of the minerals. In this case the differences between the helium content of old and young beryls could not be so marked; to explain them it must be assumed that at the time of formation of younger beryls much of this element had disintegrated. (The only other explanation why the younger beryls contained so little helium would be the assumption, for geological reasons, that less of this hypothetical element was included during their formation; but there does not seem to be any independent geological evidence in favour of a different origin of the younger and older beryls.)

---

* These considerations are quite independent of the attempts to observe ^{8}Be by means of positive rays. If ^{8}Be is present at all, its quantity is certainly less than 1×10^{-5} g./g. beryllium (Nier 1937).

By plotting the maximum, or the median, helium contents of beryls against the presumed time of formation of the minerals, a rather steep curve results, the nearest theoretical approximation to it is gained on the assumption of a half-value period of the hypothetical helium-producing element of about $1{\cdot}5 \times 10^8$ years. If this be of the right order, the helium production in beryls to-day can only be about $4–8 \times 10^{-12}$ c.c./g. beryl/year. This is hardly sufficient for a direct experimental test.

The best methods of attacking this problem of the helium content of beryls at present seem

(1) To analyse chemically, as completely as possible, beryls of the same geological age and a different helium content; perhaps differences in the content of rarer elements will be suggestive.

(2) To test the various chemical elements present in beryls in order to find out whether they generate helium.

Attention should be directed in the first place to the lighter elements, since it is to be presumed that in spite of the relatively short half-value period no detectable rays are emitted.

The table of isotopes may seem, on mass-energy considerations, to exclude the possibility of a spontaneous disintegration, but in the case of other elements the position is exactly the same as with beryllium: there may be unknown isotopes in very small quantities which decayed, or are still decaying, with helium formation, but without emission of detectable rays. Obviously, the same experimental method of attack as has been used for beryllium is available for the problem, and as a preliminary step in this direction the elements lithium, sodium and potassium—which are all present in beryls—have been investigated. The results are given below.

In the case of potassium the above figures confirm previous experiments in which an upper limit of $0{\cdot}4 \times 10^{-11}$ c.c. of helium/g./year was found

(Paneth and Peters 1928). None of the elements investigated produce at present sufficient helium to explain the helium content of the beryls. Considerations similar to those described above can be applied to prove that even a formerly stronger helium production seems hardly compatible with the facts; but further discussion is postponed until more experimental material is available.

TABLE II. HELIUM IN VARIOUS METALS

Metal	Minimum age, in years	Quantity, in g.	Minimum age × g.	Spectrum	Upper limit of He, in 10^{-10} c.c.	Upper limit of He/g./year, in 10^{-11} c.c.
Li	16	0·7	11	Ne only	3	2·7
Li	9	3·8	34	Bands	5	1·5
Be	3	7·1	21	—	1	0·5
Na	2–3	21	53	—	10	1·9
K	8	21	168	Ne; bands	5	0·3
K	30	4·5	135	Ne + He	10	0·7

To summarize, while we are unable at present to attribute the helium in beryls to any particular source it may fairly be claimed that it has been shown that, at any rate, the beryllium itself is not responsible

Besides the institutions and persons mentioned at the end of the preceding paper our thanks are due to Messrs Hopkins and Williams, Ltd., London, for the loan of several kg. of beryllium nitrate, to Professor E. Wiberg, Karlsruhe; to the National Physical Laboratory, Teddington; to the Compagnie de Produits Chimiques et Electrométallurgiques, Paris; to Degussa, Frankfurt/Main; and to Heräus Vacuumschmelze, Hanau/Main, for the gift of specimens of beryllium metal of various ages, to Professors A. Franke and H. Mark, Vienna, for old lithium samples; and to Professor A. Holmes, Durham, for information regarding geological questions.

SUMMARY

Various specimens of old beryllium metal have been analysed for helium. Since the very sensitive method used failed to detect any traces of helium it must be concluded that the spontaneous production of helium in beryllium is less than $1\cdot3 \times 10^{-11}$ c.c. of helium/g. beryllium/year.

From this figure it follows that the helium content of beryls cannot be explained as a consequence of the spontaneous disintegration of a beryllium isotope of mass 8. Even the assumption that such a beryllium isotope was

present in previous geological periods and has now mostly decayed is not compatible with the very low limit found for present-day helium production.

In recent years the helium content of beryls has been attributed to the influence of γ-rays from radioactive minerals in the neighbourhood of the beryls, and to cosmic radiation. From the figures, however, given in the preceding paper about the amount of helium produced in beryllium by γ-rays it follows that the influence of the natural sources of γ-radiation is not nearly sufficient to explain the helium content of beryls.

Since, therefore, beryllium does not produce adequate amounts of helium, either under the influence of external radiation or as a consequence of spontaneous disintegration, it seems that the helium content of beryls is not connected with its beryllium content at all, but is due to some other chemical element.

References

Atkinson and Houtermans 1929 *Nature, Lond*, **123**, 567.
Bernardini and Mando 1935 *Phys. Rev.* **48**, 468.
Born 1936 *Naturwissenschaften*, **24**, 73.
Burkser, Kapustin and Kondogury 1937 *C.R., Moscow (Doklady)*, **15**, 193.
Chadwick and Goldhaber 1935 *Proc. Roy. Soc.* A, **151**, 479.
Evans and Henderson 1933 *Phys. Rev.* **44**, 59.
Fränz and Steudel 1934 *Phys. Z.* **35**, 219.
Friedländer 1935 *C.R. Acad. Sci., Paris*, **201**, 337.
Gans, Harkins and Newson 1933 *Phys. Rev.* **44**, 310.
Gluckauf and Paneth 1938 *Proc. Roy. Soc.* A, **165**, 229.
Hahn 1934 *Naturwissenschaften*, **22**, 744.
— 1935 *Forsch. Fortsch dtsch. Wiss.* **11**, 424.
— 1936 *Vernadsky Festschrift (Moscow)*, p. 485.
Langer and Raitt 1933 *Phys. Rev.* **43**, 585.
Libby 1933 *Phys. Rev.* **44**, 512.
Nier 1937 *Phys. Rev.* **52**, 936.
Paneth and Peters 1928 *Z. phys. Chem.* B, **1**, 170.
Paneth, Gluckauf and Loleit 1936 *Proc. Roy. Soc* A, **157**, 412
Rayleigh 1929 *Nature, Lond.*, **123**, 607.
——— 1933 *a* *Nature, Lond.*, **131**, 724.
——— 1933 *b* *Proc. Roy. Soc.* A, **142**, 370.
Strutt 1908 *Proc. Roy Soc.* A, **80**, 572.
— 1910 *Proc. Roy. Soc.* A, **84**, 194.
Szilard and Chalmers 1934 *Nature, Lond.*, **134**, 494.

On the neutrino theory of light

By M. H. L. Pryce, *Trinity College, Cambridge*

(*Communicated by P. A. M. Dirac, F.R.S.—Received* 21 *December* 1937)

1 Introduction

The neutrino theory of light, developed by Jordan, Kronig and others, is based on an idea of de Broglie's, who suggested that one should regard a photon, not as a simple system, but as composed of two elementary particles bound together in some way (de Broglie 1932, 1933, 1934). According to de Broglie these particles are to be regarded as complementary in the same sense as electrons and positrons are complementary. This idea gives a good account of the creation and annihilation of photons and allows them to obey the Bose-Einstein statistics while retaining the idea that all elementary particles obey the Fermi-Dirac statistics and possess spin angular momentum $\frac{1}{2}\hbar$.

As developed by de Broglie, however, the hypothesis led to the consequence that the component particles each had exactly equal energy and momentum; thus if the state of the photon is known the state of the components can be inferred. The fact that, although the representatives are symmetrical with respect to interchange of two photons, it is then nevertheless impossible for two photons to be in the same state, owing to the underlying Fermi-Dirac statistics of the components, and so the photons would to all intents and purposes themselves obey the Fermi-Dirac statistics, led Jordan to modify the hypothesis (Jordan 1935) He suggested that it is not the interaction between the neutrinos and antineutrinos that binds them together into photons, but rather *the manner in which they interact with charged particles that leads to the simplified description of light in terms of photons*. To account for the process usually interpreted as the emission of a photon of circular frequency ω from an atom, he proposed that one of the following two processes can occur

(*a*) A neutrino, or antineutrino, passing through the atom receives an impulse, thereby gaining energy $\hbar\omega$ and momentum $\hbar\omega/c$, the atom at the same time losing that amount of energy and momentum.

(*b*) The atom emits simultaneously a neutrino with energy E', momentum p', and an antineutrino with energy $\hbar\omega - E'$, momentum $\hbar\omega/c - p'$.

It must be stressed, however, that this formulation of (*b*) is unduly loose; the neutrino and antineutrino are not really ejected in a state in which the energy and momentum of each is definite, but into a superposition of such states with quite definite weights and phases.

Although in his original formulation Jordan did not distinguish between neutrinos and antineutrinos, the distinction was in fact implicit in his mathematics. Born and Nagendra Nath clarified this point and showed that the mathematical analysis gains considerably in elegance if an antineutrino is described as a hole in a state of negative energy of the neutrinos (Born and Nagendra Nath 1936*a*). Such a description is purely a mathematical convenience and entails no physical assumption beyond the fact that neutrinos and antineutrinos are created and annihilated together in pairs. Conceptually, too, this description has advantages, for it reduces process (*b*) to a special case of (*a*) in which the neutrino is originally in a state of negative energy and jumps to one of positive energy. According to the new hypothesis, then, a photon is to be associated, not with a bound pair of particles, but with a transition of a neutrino from one state to another. This applies both to the emission and the absorption of a photon.

In order to agree with observations we are led to ascribe the following properties to the hypothetical neutrinos, which we write as postulates:

I. The rest-mass of the neutrino is zero.

II. The direction of motion of a neutrino is unchanged by its "radiative" interaction with charged particles. (This does not mean that other types of interaction, such as the one involved in β-decay, cannot deflect the neutrino.)

III. Neutrinos obey the Fermi-Dirac statistics.

IV. The neutrino possesses spin angular momentum $\frac{1}{2}\hbar$.

We are led to I and II by a consideration of the balance of energy and momentum in a radiative process; for if the neutrino gains energy $\hbar\omega$ it must gain momentum $\hbar\omega/c$, this being the observed relation between recoil momentum and energy of the atom. Writing $\mathbf{p}$, $\mathbf{p}'$ for the momentum before and after, and μ for the rest-mass, this implies

$$|\mathbf{p}'-\mathbf{p}| = \hbar\omega/c,$$

$$\sqrt{(\mu^2c^2+\mathbf{p}'^2)}-\sqrt{(\mu^2c^2+\mathbf{p}^2)} = \hbar\omega/c.$$

From this it is easy to derive

$$[(\mu\hbar\omega)^2]+[\mathbf{p}^2\mathbf{p}'^2-(\mathbf{p}\,.\,\mathbf{p}')^2] = 0.$$

Neither of the square brackets can be negative; they must therefore both vanish. This implies that $\mu = 0$, and $\mathbf{p}$ and $\mathbf{p}'$ are parallel. If the momentum of the neutrino has the value $\mathbf{p}$ its energy can thus have the two values $\pm c\,|\mathbf{p}|$. The unit vector $c\mathbf{p}/H$ defines the direction of motion of the neutrino; if H is negative, $\mathbf{p}$ is opposite to the direction of motion. A closer analysis shows that $c\mathbf{p}/H$ and $c\mathbf{p}'/H'$ cannot be antiparallel.

Another argument for postulate I is to be derived from the fact that light *in vacuo* is propagated with the velocity c, independent of frequency or other circumstance, and therefore the neutrinos giving rise to this effect must travel with velocity c.

We might perhaps object that the two above laws are not really established with universal validity. For instance, we might suppose that light does not really travel with velocity c but that our instruments are too gross to detect the difference. In such a case we can at least establish an upper limit for μ; for if a star moves with a transverse velocity relative to the earth neutrinos travelling from it with differing velocities will arrive into a telescope from different directions, if μ is not zero, red light will on the whole be associated with slower moving neutrinos than blue light, and the star would appear as a spectral band instead of a point. Allowing a velocity of 100 km./sec. and a resolving power of 1 sec. for visible light, this sets a limit of $\mu/m < 10^{-6}$ for the ratio of μ to the mass of the electron. Another limit can be obtained by considering the condition that the two particles emitted in process (b) from a star should arrive on earth sufficiently simultaneously to appear as a photon of frequency ω; for were this not the case no sharp spectral lines would be observed from the star. A train of waves associated with a photon usually has a length of about a metre (i e. the life-time of atomic states is of the order of 10^{-8} sec.); for a spectral line to be observable the neutrino and antineutrino must not have separated by more than a metre in their journey through space, which can be of the order of 10^{27} cm. This sets a limit of $\mu/m < 10^{-17}$. These arguments, of course, have no bearing on the neutrino involved in β-decay.

Similar empirical arguments may be brought forward in support of postulate II. For, as Scherzer has pointed out, if the two directions differ by a finite angle, although the general intensity falls off as the inverse square of the distance, the intensity of a *spectral line* falls off as the inverse fourth power if the angle of divergence is greater than the angle subtended by the spectroscope at the source; this could not well be reconciled with astronomical observations (Scherzer 1935).

Postulates III and IV are not based on any observations of the properties of light, but may rather be said to be the *raison d'être* of the theory.

It must be borne in mind that the particles in process (b) are *emitted* exactly in the same direction and no interaction is required to keep them together, for no interaction (except a β-disintegration) can ever change this. This exact equality gives to the theory a rather artificial look. It may not seem so artificial, however, if one states postulate II in the following way: The interaction energy commutes with the direction of motion of the neutrino; for it is quite simple to construct observables of this type (cf. equation (2·6)).

On such a hypothesis the polarization properties of light must in some way be connected with the spin of the neutrino Without going into the question at all deeply we can deduce a few features of the connexion. It is well known that the angular momentum of a photon about its direction of motion can have the two eigenvalues $\pm\hbar$; this means that the angular momentum of the atom about this direction changes by $\hbar$ or $-\hbar$. On the neutrino theory of light this states that the neutrino always makes a transition in which the component of spin in the direction of motion is reversed, i.e. changes from $\pm\frac{1}{2}\hbar$ to $\mp\frac{1}{2}\hbar$.* We shall later derive this fact rigorously (equations (6·3) and (6·4)). If the original state is one of negative energy, corresponding to (*b*), the consequent antineutrino has spin in the opposite direction from the corresponding negative energy neutrino; the neutrino and antineutrino therefore have parallel spins. If the transition is from an *eigenstate* of spin in the forward direction to one in the backward direction, the light will be circularly polarized in the right-hand direction, and conversely; if the transition is from a superposition of two states of opposite spin, with equal weights, to the corresponding opposite state, the light will be plane-polarized.

The central problem of the theory is to find a law of interaction between neutrinos and charged particles which will reproduce the observed features of radiation. Fortunately this is not so formidable a task as would at first appear, for these features can be described completely if certain observables, satisfying known commutation rules, are known. These observables make their appearance both when radiation is described in terms of waves and electromagnetic fields, and when it is described in terms of photons; in the first case they appear as the Fourier amplitudes of the field strengths, and in the second as the quantized amplitudes of the photon states when the method of second quantization is applied to the photons. Both the Bose-Einstein statistics of the photons and Maxwell's equations follow as a consequence of their commutation rules and the equations giving their rate of change with time, and further, if they are known, the interaction energy of matter and radiation can be written down immediately. The problem therefore reduces to finding an expression for these amplitudes in terms of the quantities describing the assembly of neutrinos, or, as we shall say, the neutrino field. If we can do this, not only are we certain that the phenomena of radiation will be described by the theory, but by writing down the interaction Hamiltonian we can discover how neutrinos and charged particles interact.

---

* That this is violated in Kronig's 1936 paper is closely connected with the lack of invariance of his theory under rotation; as always in quantum theory, transformation under rotation and angular momentum are closely connected.

The most suitable way of describing the neutrino field is by means of the amplitudes which play a similar role to the radiation amplitudes, but satisfy a different set of commutation relations, characteristic of the Fermi-Dirac statistics. Jordan considered that the essential part of the problem was to construct Bose-Einstein amplitudes from Fermi-Dirac amplitudes, and that in this connexion the spin and polarization were inessential complications. He therefore studied a one-dimensional model in which neutrinos were supposed to have no spin and photons no polarization. The neutrinos all travel in the positive direction, and to each value of the momentum (ranging from zero to infinity) there correspond two amplitudes. (In his original papers it was not made clear whether the two amplitudes correspond to spin (as Kronig (1935) assumed in his developments), or to two kinds of particles.) The photons also travel in the positive direction and are described by one amplitude for each value of the momentum. Jordan found an expression for the photon amplitudes in terms of the neutrino amplitudes which varied correctly with time and satisfied the correct commutation rules, but his proof of the latter depended on certain assumptions of convergence, which, although justified, were not explicitly stated; a more naïve proof would in fact have led to expressions of the form $\infty-\infty$. Later developments by Jordan (1936*a*, *b*) and by Born and Nagendra Nath (1936 *a*) led to simplification of the formulation and clearer understanding of the convergence conditions. The best formulation is that of Born and Nagendra Nath, who use the idea of holes, whereby only one amplitude is necessary to describe the neutrino field and the momentum varies from $-\infty$ to ∞, and state the convergence conditions clearly; these conditions amount to saying that there is only a finite number of neutrinos in states of positive energy and only a finite number *absent* from states of negative energy; i.e there is only a finite number of physically observable particles. How necessary the convergence conditions are is shown by the fact that Fock (1937*a*, *b*), through disregarding them, was led to believe the theory untenable. Recently, a very clear paper by Sokolow (1937) has thrown considerable light on Jordan's original formulation; Sokolow shows that the two amplitudes used there are closely connected with the use of a matrix representation

$$\alpha = \begin{pmatrix} 0, & 1 \\ 1, & 0 \end{pmatrix},$$

where α is the quantity occurring in the "one-dimensional Dirac wave equation"

$$H\psi = c\alpha p\psi,$$

H being the energy and p the momentum. The analysis of Born and Nagendra Nath corresponds to the more natural choice

$$\alpha = \begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}.$$

In the usual theory the energy of the radiation is given by a simple expression in terms of the amplitudes. The rate of change of all quantities describing the field is given by their commutator with this energy. According to the neutrino hypothesis, however, the rate of change of any quantity is its commutator with the total *neutrino energy*. Thus the commutator of all the photon amplitudes with the difference of these two energies must be zero; i.e. the difference commutes with all the photon amplitudes. Kronig, in a series of papers (Kronig 1935 *a*, *b*, *c*), studying the number of ways in which the same state of the radiation field could be realized from the neutrino field, found a very simple expression for this difference; neglecting certain inessential terms which can be avoided by a trick (Jordan 1936 *a*), it is proportional to the square of B, the number of neutrinos in excess over the number of antineutrinos; or, in terms due to Jordan, the square of the total neutrino charge.

It is easy to calculate the wave function which on the one-dimensional model describes the pair of particles produced in process (*b*) when an atom jumps to a lower state, emitting energy $\hbar\omega$. Writing x, y for the co-ordinates of the neutrino and antineutrino respectively it is

$$\psi = \frac{e^{i\omega x/c} - e^{i\omega y/c}}{i(x-y)}.$$

This separates when we transform to new variables: $X = \frac{1}{2}(x+y)$, the co-ordinate of the "centre of gravity", and $r = \frac{1}{2}(x-y)$, the half-separation:

$$\psi = \frac{\sin \omega r/c}{r} e^{i\omega X/c}.$$

Thus the centre moves with momentum $\hbar\omega/c$, while the inner state is described by the wave function $(\sin \omega r/c)/r$. Such a state of affairs is not too dissimilar to the one conceived by de Broglie.

The passage from the one-dimensional model is discussed in a joint paper by Jordan and Kronig (1936), who show that no additional difficulties arise with the statistics, but do not go into the question of polarization. They show how Planck's law follows; the interesting point emerges from their work that the neutrinos can never attain thermal equilibrium, this is not very mysterious, for they possess other integrals of motion besides the total energy, such as the individual directions of motion and the quantity B, and therefore an assembly of neutrinos interacting with matter does not constitute an ergodic system.

An attempt at a neutrino theory of light giving the polarization was made by Scherzer (1935), but his neutrinos, instead of possessing spin $\frac{1}{2}\hbar$, were transversely polarized like photons; furthermore, his theory did not give the correct statistics for the photons.

The complete success of the neutrino hypothesis seemed to be established in a recent paper by Kronig (1936), in which he gives an expression for the electromagnetic field strengths in terms of the neutrino amplitudes, from which the questions concerning polarization can be answered. Unfortunately owing to an oversight, the theory is not invariant under a rotation of the co-ordinate system.

The present paper studies the conditions imposed by the commutation rules of the amplitudes and those imposed by the connexion between spin and polarization; *the result of the investigation is that these conditions are mutually incompatible* In so far as the failure of the theory can be traced to any one cause it is fair to say that it lies in the fact that light waves are polarized transversely while neutrino "waves" are polarized longitudinally, and for "group-theoretical" reasons it is impossible to construct the former from the latter in an invariant manner. It is *not* the case that no law can be found which could reproduce correctly the phenomena of radiation, but that there is an embarrassingly infinite number of such laws, violating the law of conservation of angular momentum for the neutrinos, it is true, all of which are equally entitled to selection, so that, according to the principle of relativity, none may be singled out as being the correct one *

Unless some fundamental modification is made, therefore, Jordan's hypothesis must be abandoned. It is to be hoped that the really beautiful mathematical theory which has been developed in the course of its three years of life may eventually find application somewhere in physics.

---

* Kronig's theory is of course a particular case of one of these laws.

2. QUANTUM MECHANICS OF THE NEUTRINO

We shall try to make as few assumptions as possible about the properties of neutrinos. We shall assume that there is no direct coupling between the neutrinos; that is to say, the energy, momentum, angular momentum, of a collection of neutrinos is the sum of the separate energies, momenta and angular momenta; we shall also make use of the assumption in a rather stronger form than this later (p. 265). Further, we assume that in a state where the momentum has the value $\mathbf{p}$, the possible values of the energy are $\pm c\,|\,\mathbf{p}\,|$ (postulate I); this is embodied in the equation

$$H^2 = c^2\mathbf{p}^2. \tag{2·1}$$

We shall replace postulate IV by a weaker one, namely, that the component of angular momentum of a neutrino in the direction of its motion (this can be formulated precisely as we shall see in a moment) has the two eingenvalues $\pm \frac{1}{2}\hbar$. This means that we need not assume the position of the neutrino to be observable nor that the angular momentum is the sum of an orbital angular momentum ($[\mathbf{q}\times\mathbf{p}]$, where $\mathbf{q}$ is the position observable) and the spin $\frac{1}{2}\hbar\boldsymbol{\sigma}$ (σ_1, σ_2, σ_3 being observables capable of representation by Pauli matrices).

From considerations of the invariance of physical laws under translations in time and space, rotations in space, and Lorentz transformations, it follows that there exist ten fundamental displacement observables, satisfying the commutation relations of the ten infinitesimal operators of the heterogeneous Lorentz group (apart from a factor $i\hbar$). These are: the energy, the three components of linear momentum, the three components of angular momentum and the three components of a vector, which, owing to the fact that it contains the time variable explicitly and usually plays a less important role than the others, has no special name. Regarded from the standpoint of a four-dimensional continuum these quantities are the four components of a vector and the six components of an antisymmetrical tensor respectively. Let us denote them by H, $\mathbf{p}$, $\mathbf{M}$ and $\mathbf{N}$.* Using Greek suffixes to denote the components of a vector, and writing $\epsilon_{\alpha\beta\gamma}$ for the completely antisymmetrical array of third rank, whose components are

---

* $\mathbf{N}+ct\mathbf{p}$ does not contain the time explicitly; c times this may conveniently be called the moment of energy; that this is a reasonable name may be seen by looking at the form of $\mathbf{N}+ct\mathbf{p}$ for special systems, as for instance the electromagnetic field, where it is $(\frac{1}{8}\pi c)\int \mathbf{x}(\mathbf{E}^2+\mathbf{H}^2)\,dxdydz$, $\mathbf{x}$ being the vector with components (x, y, z).

$+1$, -1 or 0 according as $\alpha\beta\gamma$ are an even, an odd or no permutation of 123, we can write the commutation rules of these quantities as follows[*]

$$\begin{aligned} [M_\alpha, p_\beta] &= \epsilon_{\alpha\beta\gamma} p_\gamma & [N_\alpha, H] &= c p_\alpha, \\ [M_\alpha, M_\beta] &= \epsilon_{\alpha\beta\gamma} M_\gamma & [N_\alpha, p_\beta] &= \delta_{\alpha\beta} H/c, \\ [M_\alpha, N_\beta] &= \epsilon_{\alpha\beta\gamma} N_\gamma & [N_\alpha, N_\beta] &= -\epsilon_{\alpha\beta\gamma} M_\gamma. \end{aligned} \tag{2·2}$$

The brackets of the other pairs, not mentioned here, vanish.[†]

We can define the direction of motion of a neutrino by the unit vector $\mathbf{n}$, formed as follows

$$\mathbf{n} = c\mathbf{p}H^{-1} = cH^{-1}\mathbf{p}. \tag{2·3}$$

The components of $\mathbf{n}$ are real[‡] observables because $\mathbf{p}$ and H commute, $\mathbf{n}$ is a unit vector because of equation (2·1). As we have already said, the direction of motion is opposite to the momentum if the energy is negative.

From (2·2) we see that $\quad \mathbf{M}.\mathbf{p} = \mathbf{p}.\mathbf{M},$

and therefore this is a real observable. Also

$$[\mathbf{M}.\mathbf{p}, H] = 0,$$

$$[\mathbf{M}.\mathbf{p}, \mathbf{p}] = 0.$$

We can therefore define the component of angular momentum in the direction of motion to be $\mathbf{M}.\mathbf{n} = c(\mathbf{M}.\mathbf{p})H^{-1}$, it is a real observable and commutes with $H, \mathbf{p}, \mathbf{n}, \mathbf{M}$. Let us call it $\frac{1}{2}\hbar\gamma$ Our modified postulate IV then becomes

$$\mathbf{M}.\mathbf{n} = \tfrac{1}{2}\hbar\gamma, \tag{2·4}$$

$$\gamma^2 = 1. \tag{2·5}$$

We shall need to use a representation in which H, $\mathbf{n}$ and γ are diagonal. The representative of a state may be written $(H'; \gamma'; \mathbf{n}'|)$.[§] The eigenvalues $\mathbf{n}'$ of $\mathbf{n}$ may be specified by polar angles ϑ, ϕ. In order to define the representation completely we must fix the relative phases of the basic states. It will

---

[*] The bracket of two quantities is defined by the relation $i\hbar[A, B] = AB - BA$.

[†] If we define the six-vector M_{ij} by $(M_{23}, M_{31}, M_{12}) = (M_1, M_2, M_3)$, $(M_{41}, M_{42}, M_{43}) = c(N_1, N_2, N_3)$ and the vector p_i by $(p_1, p_2, p_3, p_4) = (p_1, p_2, p_3, H)$, these commutation rules can be written much more compactly in the relativistic form

$$[M_{ij}, p_k] = g_{jk} p_i - g_{ik} p_j,$$

$$[M_{ij}, M_{kl}] = g_{il} M_{jk} - g_{ik} M_{jl} + g_{jk} M_{il} - g_{jl} M_{ik},$$

where $g_{11} = g_{22} = g_{33} = -g_{44}/c^2 = -1$; $g_{ij} = 0$, $i \neq j$.

[‡] I.e. their representatives are Hermitian.

[§] Cf. Dirac (1930, Chap. v).

not be necessary to do this for states with different values of $\mathbf{n}'$, however, for, in accordance with postulate II, we shall be interested only in observables which commute with $\mathbf{n}$, and therefore whose representatives are not affected by a change of the relative phases of states with different $\mathbf{n}'$. In order to specify the relative phases of states with different H' we introduce the real observable ρ, defined by

$$\begin{aligned} \rho &= H^{-1}(c\mathbf{N}\,.\,\mathbf{p}-\tfrac{3}{2}i\hbar H)\,H^{-1}+t \\ &= H^{-1}\mathbf{N}\,.\,\mathbf{n}-\tfrac{3}{2}i\hbar H^{-1}+t. \end{aligned} \tag{2·6}$$

It does not contain t explicitly, for the term t in its definition just counteracts the explicit dependence of $\mathbf{N}$ on t. Now $\mathbf{N}\,.\,\mathbf{p}$ is easily seen to commute with $\mathbf{n}$ and with $\mathbf{M}$. It follows that ρ commutes with $\mathbf{n}$ and $\mathbf{M}$, and therefore with γ. It has the important property of being formally conjugate to H:

$$[\rho, H] = 1. \tag{2·7}$$

The existence of such an observable is characteristic for systems whose rest mass is zero.

We can therefore choose the representation in such a way that ρ is represented by the operation $i\hbar\,\partial/\partial H'$.

$$(H';\gamma';\mathbf{n}'\,|\,\rho\psi) = i\hbar\frac{\partial}{\partial H'}(H';\gamma';\mathbf{n}'\,|\,\psi). \tag{2·8}$$

The relative phases of all states with the same $\mathbf{n}'$ will now be completely fixed if we fix the relative phases of any one pair of states with the same H' but different γ'(γ' takes the two values 1 and -1). Let us for the moment assume this to have been done. Then the matrix with components

$$(H';\gamma';\mathbf{n}'\,|\,\epsilon\,|\,H'';\gamma'';\mathbf{n}'') = \delta_{H'H''}\,\delta_{\mathbf{n}'\mathbf{n}''}\,\epsilon_{\gamma'\gamma''}, \tag{2·9}$$

where $\epsilon_{\gamma'\gamma''}$ are the components of the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

defines an observable, of which it is the representative. If we take another choice of the relative phase, say one in which the states with $\gamma' = 1$ are multiplied by a factor $e^{i\theta}$ and those with $\gamma' = -1$ are multiplied by $e^{-i\theta}$, then we get a different observable corresponding to this matrix. In the first representation this observable is represented by $\delta_{H'H''}\,\delta_{n'n''}\,\bar{\epsilon}_{\gamma'\gamma''}$, where $\bar{\epsilon}_{\gamma'\gamma''}$ is given by

$$\begin{pmatrix} 0 & e^{-i\theta} \\ e^{i\theta} & 0 \end{pmatrix}.$$

Accordingly we have a one-parameter family of observables, depending on the parameter θ, and we can fix the phases by requiring that any one of them should be represented by (2·9). Let the one chosen for this purpose be called ϵ_a.

From (2·9) we see that ϵ_a commutes with $\mathbf{n}$, H and ρ, and therefore also with $\mathbf{p}$, that it anticommutes with γ and that

$$\epsilon_a^2 = 1. \tag{2·10}$$

In order to discover the physical nature of the observable ϵ_a we shall enquire how it transforms when a change is made in the frame of co-ordinates. Since it commutes with H and $\mathbf{p}$ it is unaffected by a translation in space and time. On the other hand, it has no very simple commutation laws with the Cartesian components of $\mathbf{M}$ (in fact, we cannot write them down until we fix the relative phases of states with different $\mathbf{n}'$), which indicates that it does not transform in any simple manner under rotation Nevertheless, if we consider all states in which $\mathbf{n}$ has a definite value $\mathbf{n}'$, and rotate the frame of reference about this direction, the transformation will be determined by the component of $\mathbf{M}$ in that direction, which for these particular states is just $\mathbf{M}\,.\,\mathbf{n} = \frac{1}{2}h\gamma$. A rotation of the frame of reference through an angle θ will induce a change in ϵ_a as follows ·

$$\epsilon_a \rightarrow \tilde{\epsilon}_a = e^{-\frac{1}{2}i\theta\gamma}\epsilon_a e^{\frac{1}{2}i\theta\gamma}. \tag{2·11}$$

From (2·9) it follows that the representative of $\tilde{\epsilon}_a$ is $\delta_{H'H''}\delta_{n'n''}\tilde{\epsilon}_{\gamma'\gamma''}$, where $\tilde{\epsilon}_{\gamma'\gamma''}$ is given by

$$\begin{pmatrix} 0 & e^{-i\theta} \\ e^{i\theta} & 0 \end{pmatrix} = \cos\theta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \sin\theta\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

This can be written $$\tilde{\epsilon}_a = \cos\theta\epsilon_a + \sin\theta\epsilon_b,$$

where $\epsilon_b = -i\gamma\epsilon_a$.

It shows that ϵ_a and ϵ_b transform like the components of a vector in two mutually perpendicular directions, both perpendicular to $\mathbf{n}$. Thus ϵ_a is the component in some direction given by a unit vector $\mathbf{a}$, perpendicular to $\mathbf{n}$, of an observable vector $\boldsymbol{\epsilon}$.

In order to fix the representation, therefore, we must decide on a definite $\mathbf{a}$. Having made this choice in some way for each $\mathbf{n}$, we take $(\mathbf{a}\,.\,\boldsymbol{\epsilon})$ for the observable ϵ_a in (2·9). *This choice is entirely arbitrary, for among all unit vectors perpendicular to a given direction in space all are equivalent and none is singled out in any way.*

A rotation of $\mathbf{a}$ through an angle θ about $\mathbf{n}'$ changes the relative phase by θ.

It is this arbitrariness that is the stumbling block for the neutrino theory of light, for we must impose the restriction that the results of the theory are independent of the choice of **a**

3. DISCUSSION OF A SPECIAL CASE

In order to illustrate the results of the preceding section we shall apply them to particles described by a Dirac equation with the rest-mass term put to zero We therefore assume that the neutrinos have an observable position **q** and their Hamiltonian is given by

$$H/c = i\gamma_4(\gamma_1 p_1+\gamma_2 p_2+\gamma_3 p_3),$$

where γ_1, γ_2, γ_3, γ_4 are real observables satisfying

$$\gamma_r\gamma_s+\gamma_s\gamma_r = 2\delta_{rs} \quad (r,\, s = 1,\, 2,\, 3,\, 4).$$

This is the case actually studied by Kronig (1936).

Let us introduce the notation

$$\beta = \gamma_4,$$

$$\sigma_\alpha = -i\gamma_1\gamma_2\gamma_3\gamma_\alpha \quad (\alpha = 1,\, 2,\, 3),$$

$$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3).$$

Then it is well known that the other fundamental observables describing such a system are

$$\mathbf{M} = [\mathbf{q}\times\mathbf{p}]+\tfrac{1}{2}\hbar\boldsymbol{\sigma},$$

$$\mathbf{N} = \frac{1}{2c}\{H\mathbf{q}+\mathbf{q}H\}-ct\mathbf{p}.$$

From (2·4) we find $\qquad \gamma = \boldsymbol{\sigma}\,.\,\mathbf{n} = -\gamma_1\gamma_2\gamma_3\gamma_4.$

It commutes with **N**; for this case, therefore, γ is a relativistic invariant. The observable ρ is given by

$$\rho = \frac{1}{c}(\mathbf{n}\,.\,\mathbf{q})+\frac{i\hbar}{H} = \frac{1}{c}(\mathbf{q}\,.\,\mathbf{n})-\frac{i\hbar}{H}.$$

If an observable which is a function of **p**, **q** and the γ's commutes with H, **n** and ρ it can easily be proved that it is (1) independent of **q**, (2) homogeneous of degree zero in **p**. If, further, it anticommutes with γ, a simple calculation shows that it must be a linear combination of the components of the vector $[\mathbf{n}\times\beta\boldsymbol{\sigma}]$, the coefficients being homogeneous functions of **p**. Such an observable can be taken for the ϵ_a if in addition its square is unity. Now let $\mathbf{a}$ be a unit vector perpendicular to $\mathbf{n}$, chosen in some way for each $\mathbf{n}$. According to quantal ideas of functions of observables, $\mathbf{a}$ is a function of $\mathbf{n}$, and commutes with all observables that commute with $\mathbf{n}$. Let $\mathbf{b}$ be the unit vector $[\mathbf{n}\times\mathbf{a}]$; it is perpendicular to $\mathbf{a}$ and $\mathbf{n}$. The following results can be verified without essential difficulty:

(i) $[\mathbf{n}\times\beta\boldsymbol{\sigma}] = -[\boldsymbol{\sigma}\beta\times\mathbf{n}]$, and therefore $[\mathbf{n}\times\beta\boldsymbol{\sigma}]$ is a real observable.

(ii) $[\mathbf{n}\times\beta\boldsymbol{\sigma}]$ commutes with H, $\mathbf{n}$, ρ, and therefore with $\mathbf{a}$, $\mathbf{b}$, and anticommutes with γ.

(iii) $(\mathbf{a}\,.\,\beta\boldsymbol{\sigma}) = (\beta\boldsymbol{\sigma}\,.\,\mathbf{a})$, $(\mathbf{b}\,.\,\beta\boldsymbol{\sigma}) = (\beta\boldsymbol{\sigma}\,.\,\mathbf{b})$, from which it follows that $(\mathbf{a}\,.\,\beta\boldsymbol{\sigma})$, $(\mathbf{b}\,.\,\beta\boldsymbol{\sigma})$ are real observables.

(iv)
$$(\mathbf{a}\,.\,\beta\boldsymbol{\sigma}) = (\mathbf{b}\,.\,[\mathbf{n}\times\beta\boldsymbol{\sigma}]) = -(\mathbf{b}\,.\,i\gamma\beta\boldsymbol{\sigma}),\text{[*]}$$
$$(\mathbf{b}\,.\,\beta\boldsymbol{\sigma}) = -(\mathbf{a}\,.\,[\mathbf{n}\times\beta\boldsymbol{\sigma}]) = (\mathbf{a}\,.\,i\gamma\beta\boldsymbol{\sigma}),$$

from which it follows that the components of $i\gamma\beta\boldsymbol{\sigma}$ at right angles to $\mathbf{n}$ are the same as those of $\beta\boldsymbol{\sigma}$, but rotated through a right angle about $\mathbf{n}$, and that $(\mathbf{a}\,.\,\beta\boldsymbol{\sigma})$, $(\mathbf{b}\,.\,\beta\boldsymbol{\sigma})$ commute with H, $\mathbf{n}$ and ρ, and anticommute with γ

(v)
$$(\mathbf{a}\,.\,\beta\boldsymbol{\sigma})^2 = (\mathbf{b}\,.\,\beta\boldsymbol{\sigma})^2 = 1,$$
$$(\mathbf{a}\,.\,\beta\boldsymbol{\sigma})(\mathbf{b}\,.\,\beta\boldsymbol{\sigma})+(\mathbf{b}\,.\,\beta\boldsymbol{\sigma})(\mathbf{a}\,.\,\beta\boldsymbol{\sigma}) = 0.$$

We may therefore take $(\mathbf{a}\,.\,\beta\boldsymbol{\sigma})$ for our ϵ_a.

At this point we may compare our results with Kronig's It will readily be seen that his equation (17) (Kronig 1936) is the one which fixes the relative phases of states with different γ', which he calls A and C respectively, and that taken with (19) it is equivalent to choosing

$$\mathbf{a} = \frac{[[\mathbf{n}\times\mathbf{l}]\times\mathbf{n}]}{\sqrt{[\mathbf{n}\times\mathbf{l}]^2}},$$

where $\mathbf{l}$ is the vector whose components are $(1, 1, 1)$. Nothing in nature, however, singles out this vector to play a fundamental role in the theory, and if the theory is to be satisfactory we shall have to show that the results are independent of the choice of $\mathbf{l}$. That this is not the case is most clearly shown by an example that Kronig himself has suggested.[‡] Let us consider waves in the z-direction (i.e. $\mathbf{n}' = (0, 0, 1)$). Kronig's equations then give

$$C = (\gamma_2-\gamma_1)A/\sqrt{2}.$$

If we now pass to a co-ordinate system arising by rotating the old one about the z-axis through an angle θ, the spinors A and C transform to

$$A' = SA, \quad C' = SC,$$

where
$$S = \cos\tfrac{1}{2}\theta + \gamma_1\gamma_2\sin\tfrac{1}{2}\theta.$$

---

[*] The components of $i\gamma\beta\boldsymbol{\sigma}$ are $\gamma_1, \gamma_2, \gamma_3$.

[‡] In a private communication.

From this it easily follows that

$$C' = e^{-i\theta}(\gamma_2 - \gamma_1) A'$$

and therefore the connexion between A' and C' is not the same as the connexion between A and C. If we were to follow through Kronig's construction of an electromagnetic field from the neutrino field with the new A' and C', we should get a different field from the original.

4. SECOND QUANTIZATION OF THE NEUTRINO FIELD

The formalism of the second quantization* affords an excellent mathematical apparatus for the description of processes taking place in the neutrino field. In order to avoid a representation of the neutrino states based on an observable with a continuum of eingenvalues, we shall resort to the artifice of making space periodic. This does not modify the essential results of §2, nor is it really essential in what follows We can look upon the artifice of periodicity as a restriction of the states of the system which lead to the same physical results for the points (x, y, z) and $(x+lL, y+mL, z+nL)$, l, m, n being integers and L the side of the periodicity cube. From the standpoint of displacement operators this means we are considering only the simultaneous eigenstates of $e^{iLp_1/\hbar}$, $e^{iLp_2/\hbar}$, $e^{iLp_3/\hbar}$ with the eigenvalue 1.‡ Thus the Cartesian components of momentum can only take values which are integral multiples of h/L, and the direction of n is restricted to rays with direction ratios $(\kappa_1, \kappa_2, \kappa_3,)$, $\kappa_1, \kappa_2, \kappa_3$ being relatively prime integers (including zero and negative integers). For states with a given eigendirection $\mathbf{n}'$, the energy takes the values $H' = r\hbar\omega_0$, where

$$\omega_0 = c\kappa_0 = (2\pi c/L)\sqrt{(\kappa_1^2 + \kappa_2^2 + \kappa_3^2)}.$$

Only those quantities which commute with $e^{iLp_1/\hbar}$, etc., can be considered as observables; this means that ρ is no longer an observable, but only certain periodic functions of it such as $e^{i\omega_0\rho}$. From (2·8) it follows that

$$(H'; \gamma'; \mathbf{n}' \,|\, e^{i\omega_0\rho}\psi) = (H' - \hbar\omega_0; \gamma', \mathbf{n}' \,|\, \psi).$$

Since the fundamental states are now discrete we can use the integers r to label the representatives instead of $H' = r\hbar\omega_0$. They are now $(r; \gamma'; \mathbf{n}' |)$ and the above equation reads

$$(r; \gamma'; \mathbf{n}' \,|\, e^{i\omega_0\rho}\psi) = (r-1; \gamma'; \mathbf{n}' \,|\, \psi).$$

---

* Jordan and Wigner (1928).

‡ More strictly one could allow them to have the eigenvalue $e^{i\alpha}$, where α is real. For some purposes $\alpha = \pi$ is a convenient choice (Jordan 1936 *a*); the momentum then takes on half-integral values.

This can be looked on as fixing the relative phases of states with the same $\mathbf{n}'$ and γ' but different r, instead of (2·8), which is now meaningless. The relative phases of states with different γ' are fixed as before.

Let us temporarily arrange the basic states $\psi(r; \gamma'; \mathbf{n}')$ in a sequence $\psi_1, \psi_2, \ldots$, in some order or other. Then we can form states of the neutrino field in which there are n_1 neutrinos in ψ_1, n_2 in ψ_2 and so on ($n_1, n_2, \ldots = 0$ or 1), and these will form a complete orthogonal system for the neutrino field. The phases of these states can be fixed from the phases of $\psi_1, \psi_2, \ldots$ (cf. Jordan and Wigner 1928). Let us call them $\Psi(n_1, n_2, \ldots)$. We define the amplitudes α_i and their adjoints[*] $\alpha^\dagger$, as follows

$$\begin{aligned} \alpha_i \Psi(n_1, n_2, \ldots, n_i, \ldots) &= \pm n_i \Psi(n_1, n_2, \ldots, n_i - 1, \ldots), \\ \alpha_i^\dagger \Psi(n_1, n_2, \ldots, n_i, \ldots) &= \pm (1 - n_i) \Psi(n_1, n_2, \ldots, n_i + 1, \ldots), \end{aligned} \tag{4·1}$$

where the $\pm$ depends on the $n_1, n_2, \ldots$ in some way which need not concern us here (cf. Jordan and Wigner 1928) These amplitudes obey the commutation rules

$$\left. \begin{aligned} \alpha_i \alpha_j + \alpha_j \alpha_i &= 0, \\ \alpha_i \alpha_j^\dagger + \alpha_j^\dagger \alpha_i &= \delta_{ij}. \end{aligned} \right\} \tag{4·2}$$

Also, denoting by n_i the operator which multiplies the state $\Psi(n_1, n_2, \ldots)$ by the corresponding n_i, we have

$$\alpha_i^\dagger \alpha_i = n_i. \tag{4·3}$$

To each basic state there corresponds an amplitude α and its adjoint $\alpha^\dagger$. Let us denote the amplitudes corresponding to $\psi(r, +1, \mathbf{n}')$ by $a_r^1(\mathbf{n}')$, $a_r^{1\dagger}(\mathbf{n}')$ and those corresponding to $\psi(r; -1; \mathbf{n}')$ by $a_r^2(\mathbf{n}')$, $a_r^{2\dagger}(\mathbf{n}')$. The upper index takes the values 1, 2 according as $\gamma' = +1$ or $\gamma' = -1$. The lower index refers to the value of the energy ($H' = r\hbar\omega_0$), and the appropriate direction is written inside the parentheses. The commutation rules now read

$$\begin{aligned} & a_r^\mu(\mathbf{n}') a_s^\nu(\mathbf{n}'') + a_s^\nu(\mathbf{n}'') a_r^\mu(\mathbf{n}') = 0, \\ & a_r^\mu(\mathbf{n}') a_s^{\nu\dagger}(\mathbf{n}'') + a_s^{\nu\dagger}(\mathbf{n}'') a_r^\mu(\mathbf{n}') = \delta_{rs} \delta_{\mu\nu} \delta(\mathbf{n}', \mathbf{n}''). \end{aligned} \tag{4·4}$$

The amplitudes transform under a change of basis of the neutrino states like the representatives of a state. If we change the representation by rotating the vector $\mathbf{a}$ through an angle about $\mathbf{n}'$ (by which we fix the phases in a different way), the representatives of a state change according to the law

$$\begin{aligned} (r; +1; \mathbf{n}' |) &\rightarrow (r; +1; \mathbf{n}' |)^\sim = e^{-\frac{1}{2}i\theta}(r; +1; \mathbf{n}' |), \\ (r; -1; \mathbf{n}' |) &\rightarrow (r; -1; \mathbf{n}' |)^\sim = e^{\frac{1}{2}i\theta}(r, -1; \mathbf{n}' |). \end{aligned}$$

---

[*] We here follow the terminology of the modern theory of linear operators in using the term "adjoint" instead of Dirac's "conjugate imaginary".

The amplitudes therefore transform under a rotation of **a** through an angle θ about $\mathbf{n}'$ according to the law

$$a_r^1(\mathbf{n}') \to \tilde{a}_r^1(\mathbf{n}') = e^{-\frac{1}{2}i\theta} a_r^1(\mathbf{n}'),$$

$$a_r^2(\mathbf{n}') \to \tilde{a}_r^2(\mathbf{n}') = e^{\frac{1}{2}i\theta} a_r^2(\mathbf{n}').$$

If we denote the matrix $\begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}$

by $\boldsymbol{\gamma}$ (components $\gamma_{\mu\nu}$), this can be written in the form

$$a_r^\mu(\mathbf{n}') \to \tilde{a}_r^\mu(\mathbf{n}') = \sum_{\nu=1}^{2} S_{\mu\nu} a_r^\nu(\mathbf{n}'),$$

$$a_r^{\mu\dagger}(\mathbf{n}') \to \tilde{a}_r^{\mu\dagger}(\mathbf{n}') = \sum_{\nu=1}^{2} a_r^{\nu\dagger}(\mathbf{n}') S_{\nu\mu}^\dagger,$$

$$\mathsf{S} = e^{-\frac{1}{2}i\theta\boldsymbol{\gamma}}, \quad \mathsf{S}^\dagger = e^{\frac{1}{2}i\theta\boldsymbol{\gamma}}. \tag{4·5}$$

In § 6 we shall need to consider the convergence of certain infinite series. We must remember that we are dealing with states of the neutrino field in which almost all the states of positive energy are empty and almost all those of negative energy are full. Let us denote by Ψ_N a state of the neutrino field in which all the neutrino states with a positive index r greater than N are empty and all those with a negative index r less than $-N$ are full. We shall suppose that all actually occurring states of the neutrino field can be considered as limits of sequences of such states with N tending to infinity We shall say that a series $\sum_r f_r$ converges to f if $\sum_r f_r \Psi_N$ converges to $f\Psi_N$ for all N.

We can temporarily drop the upper index from the amplitudes and also the label $(\mathbf{n}')$. Then we have

$$\left.\begin{aligned} a_r^\dagger a_r \Psi_N &= 0 \\ a_{-r}^\dagger a_{-r} \Psi_N &= \Psi_N \end{aligned}\right\} \quad r > N. \tag{4·6}$$

Multiplying these equations by a_r, $a_{-r}^\dagger$ respectively we find

$$\begin{aligned} 0 = a_r a_r^\dagger a_r \Psi_N &= a_r(1 - a_r a_r^\dagger)\Psi_N \\ &= a_r \Psi_N, \end{aligned}$$

$$0 = a_{-r}^\dagger(1 - a_{-r}^\dagger a_{-r})\Psi_N = a_r^\dagger \Psi_N,$$

i.e.

$$\left.\begin{aligned} a_r \Psi_N &= 0 \\ a_{-r}^\dagger \Psi_N &= 0 \end{aligned}\right\} \quad r > N. \tag{4·6a}$$

Any series $$\sum_{r=-\infty}^{\infty} K_r a_r^\dagger a_{r+k}$$

converges provided $k \neq 0$; for if r is sufficiently large $a_r^\dagger a_{r+k}\Psi_N$ is zero and if $-r$ is sufficiently large $a_r^\dagger a_{r+k}\Psi_N = -a_{r+k}a_r^\dagger\Psi_N$ is zero. Thus only a finite number of terms differ from zero in $\Sigma K_r a_r^\dagger a_{r+k}\Psi_N$; it therefore converges for all N.

We now calculate the sum of a very important series:

$$\sum_{r=-\infty}^{\infty}(a_r^\dagger a_r - a_{r+k}^\dagger a_{r+k}) = k. \tag{4·7}$$

A casual analysis would lead one to sum this to zero by splitting it into the two series $\Sigma a_r^\dagger a_r$, $\Sigma a_{r+k}^\dagger a_{r+k}$, which, however, both diverge. We therefore proceed as follows

$$\begin{aligned}\sum_{r=-\infty}^{\infty}(a_r^\dagger a_r - a_{r+k}^\dagger a_{r+k}) &= \lim_{M\to\infty}\sum_{r=-M}^{M}(a_r^\dagger a_r - a_{r+k}^\dagger a_{r+k})\\ &= \lim_{M\to\infty}\left\{\sum_{-M}^{-M+k-1}a_r^\dagger a_r - \sum_{M+1}^{M+k}a_r^\dagger a_r\right\}\\ &= k - \lim_{M\to\infty}\left\{\sum_{-M}^{-M+k-1}(1-a_r^\dagger a_r) - \sum_{M+1}^{M+k}(a_r^\dagger a_r)\right\}.\end{aligned}$$

According to (4·6) the second term vanishes, the sum of the series is therefore k.

A simple generalization of this series is

$$K_{\mu\nu}\sum_{r=-\infty}^{\infty}(a_r^{\mu\dagger}a_r^{\nu} - a_{r+k}^{\mu\dagger}a_{r+k}^{\nu}),$$

this being summed over repeated Greek indices. The terms with $\mu \neq \nu$ vanish, leaving

$$\begin{aligned}K_{\mu\nu}\sum_{r=-\infty}^{\infty}(a_r^{\mu\dagger}a_r^{\nu} - a_{r+k}^{\mu\dagger}a_{r+k}^{\nu}) &= K_{11}\sum_{r=-\infty}^{\infty}(a_r^{1\dagger}a_r^{1} - a_{r+k}^{1\dagger}a_{r+k}^{1})\\ &\quad + K_{22}\sum_{r=-\infty}^{\infty}(a_r^{2\dagger}a_r^{2} - a_{r+k}^{2\dagger}a_{r+k}^{2})\\ &= kK_{\mu\mu}\\ &= kT_r(\mathsf{K}).^{*}\end{aligned} \tag{4·8}$$

5. QUANTIZATION OF THE RADIATION FIELD

Let us expand the electromagnetic field strengths in terms of progressive waves. To each value of the wave vector $(\mathbf{n}')$ there correspond two independent progressive waves, which we take to be the waves polarized in the directions **a** and **b**, defined as before (pp. 257, 259):

---

* The trace $T_r(\mathsf{K})$ of the matrix K is the sum of the diagonal elements.

$$\mathbf{E} = \sum_{\mathbf{n}'} \left(\frac{\hbar\omega_0}{2L^3}\right)^{\frac{1}{2}} \sum_{k=1}^{\infty} \sqrt{(k)} \{(\xi_k \mathbf{a} + \eta_k \mathbf{b}) e^{ik\kappa_0(\mathbf{n}'.\mathbf{x}-ct)} + (\xi_k^\dagger \mathbf{a} + \eta_k^\dagger \mathbf{b}) e^{-ik\kappa_0(\mathbf{n}'.\mathbf{x}-ct)}\},$$

$$\mathbf{H} = \sum_{\mathbf{n}'} \left(\frac{\hbar\omega_0}{2L^3}\right)^{\frac{1}{2}} \sum_{k=1}^{\infty} \sqrt{(k)} \{(\xi_k \mathbf{b} - \eta_k \mathbf{a}) e^{ik\kappa_0(\mathbf{n}'.\mathbf{x}-ct)} + (\xi_k^\dagger \mathbf{b} - \eta_k^\dagger \mathbf{a}) e^{-ik\kappa_0(\mathbf{n}'.\mathbf{x}-ct)}\}. \tag{5·1}$$

The ξ_k, η_k depend on $\mathbf{n}'$ as well as on k; they satisfy the commutation relations (for the same $\mathbf{n}'$, for different $\mathbf{n}'$ they all commute)

$$\begin{gathered} \xi_k\xi_l - \xi_l\xi_k = 0, \qquad \eta_k\eta_l - \eta_l\eta_k = 0, \\ \xi_k\eta_l - \eta_l\xi_k = 0, \qquad \xi_k\eta_l^\dagger - \eta_l^\dagger\xi_k = 0, \\ \xi_k\xi_l^\dagger - \xi_l^\dagger\xi_k = \delta_{kl} \quad \eta_k\eta_l^\dagger - \eta_l^\dagger\eta_k = \delta_{kl} \end{gathered} \tag{5·2}$$

If we describe the radiation field in the language of photons the ξ_k, η_k are to be identified with the quantized amplitudes corresponding to the states of the photon with energy $k\hbar\omega_0$, momentum $k\hbar\kappa_0\mathbf{n}'$, and plane-polarized in the directions **a**, **b** respectively.* These photon amplitudes are defined as follows: Let ϕ_1, ϕ_2, ... be a complete orthogonal system of states for a photon, and form the states $\Phi(M_1, M_2, \ldots)$ of the radiation field in which there are M_1 photons in the first state, M_2 in the second and so on ($M_1, M_2, \ldots = 0, 1, 2, 3, \ldots$). These form a complete orthogonal system of states for the field. We define the amplitude b_i, and its adjoint $b_i^\dagger$ by the relations

$$\left.\begin{aligned} b_i\Phi(M_1, M_2, \ldots, M_i, \ldots) &= \sqrt{(M_i)}\Phi(M_1, M_2, \ldots, M_i - 1, \ldots), \\ b_i^\dagger\Phi(M_1, M_2, \ldots, M_i, \ldots) &= \sqrt{(M_i + 1)}\Phi(M_1, M_2, \ldots, M_i + 1, \ldots). \end{aligned}\right\} \tag{5·3}$$

They obey the commutation laws

$$\left.\begin{aligned} b_ib_j - b_jb_i &= 0, \\ b_ib_j^\dagger - b_j^\dagger b_i &= \delta_{ij}. \end{aligned}\right\} \tag{5·4}$$

Denoting by M_i the operator which multiplies $\Phi(M_1, M_2, \ldots)$ by the appropriate M_i we have the relation

$$b_i^\dagger b_i = M_i.$$

The commutation laws (5·2) are identical with (5·4) when we choose ϕ_1, ϕ_2, ... to be states of definite momentum and plane-polarization and write ξ_k, η_k for the corresponding b.

---

* Cf. Dirac (1930, Chap. XII).

If we take a different choice of the vector **a**, obtained by rotation from the original through an angle θ about **n** (cf. p. 262), the vector $\xi_k \mathbf{a} + \eta_k \mathbf{b}$ must remain the same; ξ_k, η_k therefore undergo the transformation

$$\left.\begin{aligned} \xi_k \to \tilde{\xi}_k &= \cos\theta \xi_k + \sin\theta \eta_k, \\ \eta_k \to \tilde{\eta}_k &= -\sin\theta \xi_k + \cos\theta \eta_k. \end{aligned}\right\} \tag{5·5}$$

6. THE CONNEXION BETWEEN RADIATION FIELD AND NEUTRINO FIELD

From equation (5·3) we see that ξ_k, operating on a state of the field, gives a state in which the number of photons with momentum $k\hbar\kappa_0 \mathbf{n}'$, polarization **a**, is less by one. According to Jordan's hypothesis this means a state which differs from the original by a neutrino travelling in the direction $\mathbf{n}'$ having made a transition to a state in which the momentum differs by $-k\hbar\kappa_0\mathbf{n}'$. To make this definite, suppose the original state Ψ to be one in which there are $n_r^\lambda(\mathbf{n}')$ neutrinos in each state $((r, \lambda, \mathbf{n}')\ (n_r^\lambda(\mathbf{n}') = 0 \text{ or } 1)$. Then $\xi_k \Psi$ is a linear combination of states $\Psi_{s,\mu,\nu}$ with the same occupation numbers $n_r^\lambda(\mathbf{n}')$ except for two, say $n_s^\mu(\mathbf{n}')$, $n_{s+k}^\nu(\mathbf{n}')$, which are changed to $n_s^\mu(\mathbf{n}')+1$ and $n_{s+k}^\nu(\mathbf{n}')-1$ respectively Such a state can be written (we drop the $\mathbf{n}'$)

$$\Psi_{s,\mu,\nu} = a_s^{\mu\dagger} a_{s+k}^\nu \Psi.$$

The most general expression for $\xi_k \Psi$ is therefore

$$\xi_k \Psi = \sum_{\mu,\nu,s} A_{\mu\nu}(s, s+k)\, a_s^{\mu\dagger} a_{s+k}^\nu \Psi.$$

At this point we make the assumption that the "matrix element" $A_{\mu\nu}(s, s+k)$ for the transition of a neutrino from the state $s+k$ to the state s is independent of the number of neutrinos present in other states This is a stronger form of the assumption concerning the absence of direct coupling between the neutrinos than we have heretofore used. The coefficients $A_{\mu\nu}(s, s+k)$ are then the same for all states Ψ; we can therefore write the expression for ξ_k, and similarly for η_k, in the form

$$\left.\begin{aligned} \xi_k &= \sum_{r=-\infty}^{\infty} a_r^{\mu\dagger} A_{\mu\nu}(r, r+k)\, a_{r+k}^\nu, \\ \eta_k &= \sum_{r=-\infty}^{\infty} a_r^{\mu\dagger} B_{\mu\nu}(r, r+k)\, a_{r+k}^\nu. \end{aligned}\right\} \tag{6·1}$$

Here, as in future, we sum over repeated Greek indices without indicating it explicitly.

If we take the variation with time into account, $a_r^{\mu\dagger}$ is proportional to $e^{ir\omega_0 t}$ and a_{r+k}^ν to $e^{-i(r+k)\omega_0 t}$, hence ξ_k, η_k are proportional to $e^{-ik\omega_0 t}$; this is the law of variation with time given by Maxwell's equations. The relation (6·1) will therefore lead to Maxwell's equations for the field. It will thus give a correct account of radiation phenomena if it satisfies the commutation rules (5·2) and is independent of the choice of **a**.

Equation (6·1) must still hold if we substitute in it the values of $\xi_k, \eta_k, a_r^{\mu\dagger}, a_{r+k}^{\nu}$ referred to a choice of **a** differing by a rotation through θ about **n**′ (equations (4·5) and (5·5))·

$$\cos\theta\xi_k + \sin\theta\eta_k = \sum_{r=-\infty}^{\infty} a_r^{\sigma\dagger} S_{\sigma\mu}^{\dagger} A_{\mu\nu} S_{\nu\tau} a_{r+k}^{\tau},$$

$$-\sin\theta\xi_k + \cos\theta\eta_k = \sum_{r=-\infty}^{\infty} a_r^{\sigma\dagger} S_{\sigma\mu}^{\dagger} B_{\mu\nu} S_{\nu\tau} a_{r+k}^{\tau};$$

$$\mathsf{S} = e^{-\frac{1}{2}i\theta\boldsymbol{\gamma}}, \quad \mathsf{S}^{\dagger} = e^{\frac{1}{2}i\theta\boldsymbol{\gamma}}.$$

This must hold for all values of θ, but since the rotations form a group it is sufficient if it holds for infinitesimal rotations. Differentiating with respect to θ and evaluating for $\theta = 0$, therefore, we find, after suitably renaming the dummy indices,

$$\eta_k = -\frac{i}{2}\sum_{r=-\infty}^{\infty} a_r^{\mu\dagger}\{A_{\mu\sigma}(r, r+k)\gamma_{\sigma\nu} - \gamma_{\mu\sigma}A_{\sigma\nu}(r, r+k)\}\, a_{r+k}^{\nu},$$

$$\xi_k = \frac{i}{2}\sum_{r=-\infty}^{\infty} a_r^{\mu\dagger}\{B_{\mu\sigma}(r, r+k)\gamma_{\sigma\nu} - \gamma_{\mu\sigma}B_{\sigma\nu}(r, r+k)\}\, a_{r+k}^{\nu}.$$

Comparing with (6·1) we obtain

$$\left.\begin{aligned} 2B_{\mu\nu}(r, r+k) &= -i\{A_{\mu\sigma}(r, r+k)\gamma_{\sigma\nu} - \gamma_{\mu\sigma}A_{\sigma\nu}(r, r+k)\},\\ 2A_{\mu\nu}(r, r+k) &= i\{B_{\mu\sigma}(r, r+k)\gamma_{\sigma\nu} - \gamma_{\mu\sigma}B_{\sigma\nu}(r, r+k)\}.\end{aligned}\right\} \tag{6·2}$$

This is more simply expressed in matrix notation, as on p. 262,

$$\left.\begin{aligned} 2i\mathsf{B}(r, r+k) &= \mathsf{A}(r, r+k)\boldsymbol{\gamma} - \boldsymbol{\gamma}\mathsf{A}(r, r+k),\\ -2i\mathsf{A}(r, r+k) &= \mathsf{B}(r, r+k)\boldsymbol{\gamma} - \boldsymbol{\gamma}\mathsf{B}(r, r+k).\end{aligned}\right\} \tag{6·3}$$

It follows that $\mathsf{A}(r, r+k)$ and $\mathsf{B}(r, r+k)$ have the form

$$\mathsf{A}(r, r+k) = \begin{pmatrix} 0, & u(r, r+k) \\ v(r, r+k), & 0 \end{pmatrix}, \quad \mathsf{B}(r, r+k) = \begin{pmatrix} 0, & iu(r, r+k) \\ -iv(r, r+k), & 0 \end{pmatrix},$$

the $u(r, r+k)$, $v(r, r+k)$ being numerical coefficients. The neutrinos therefore make transitions only to states of opposite spin; this agrees with our previous deductions from the balance of angular momentum.

The coefficients $u(r, s)$, $v(r, s)$ are only defined for $s > r$; we are therefore at liberty to extend the definition for $s < r$ by means of the relations

$$u(s, r) = v(r, s),$$
$$v(s, r) = u(r, s).$$

The advantage of this is that we can express A and B in terms of the u's alone

$$\mathsf{A}(r, s) = \begin{pmatrix} 0, & u(r, s) \\ \bar{u}(s, r), & 0 \end{pmatrix}, \quad \mathsf{B}(r, s) = \begin{pmatrix} 0, & iu(r, s) \\ -i\bar{u}(s, r), & 0 \end{pmatrix}. \qquad (6{\cdot}4)$$

It is also consistent with this notation to write

$$\mathsf{A}(s, r) = \mathsf{A}^\dagger(r, s),$$
$$\mathsf{B}(s, r) = \mathsf{B}^\dagger(r, s).$$

We have now to require that ξ_k, η_k should satisfy the commutation rules (5·2). Fortunately it is sufficient to consider only the equation

$$\xi_k \eta_k^\dagger - \eta_k^\dagger \xi_k = 0$$

(this is one of the equations (5·2)) in order to show that these cannot be satisfied simultaneously with (6·4). For

$$\begin{aligned}
\xi_k \eta_k^\dagger - \eta_k^\dagger \xi_k &= \sum_r \sum_s \{a_r^{\mu\dagger} A_{\mu\nu}(r, r+k)\, a_{r+k}^\nu a_{s+k}^{\tau\dagger} \bar{B}_{\sigma\tau}(s, s+k)\, a_s^\sigma \\
&\qquad - a_{s+k}^{\tau\dagger} \bar{B}_{\sigma\tau}(s, s+k)\, a_s^\sigma a_r^{\mu\dagger} A_{\mu\nu}(r, r+k)\, a_{r+k}^\nu\} \\
&= \sum_r \sum_s A_{\mu\nu}(r, r+k)\, B_{\tau\sigma}(s+k, s)\, \{a_r^{\mu\dagger} a_{r+k}^\nu a_{s+k}^{\tau\dagger} a_s^\sigma - a_{s+k}^{\tau\dagger} a_s^\sigma a_r^{\mu\dagger} a_{r+k}^\nu\} \\
&= \sum_r \sum_s A_{\mu\nu}(r, r+k)\, B_{\tau\sigma}(s+k, s)\, \delta_{rs} \{a_r^{\mu\dagger} a_r^\sigma \delta_{\nu\tau} - a_{r+k}^{\tau\dagger} a_{r+k}^\nu \delta_{\mu\sigma}\} \\
&= \sum_{r=-\infty}^{\infty} \{A_{\mu\tau}(r, r+k)\, B_{\tau\nu}(r+k, r)\, a_r^{\mu\dagger} a_r^\nu \\
&\qquad - B_{\mu\sigma}(r+k, r)\, A_{\sigma\nu}(r, r+k)\, a_{r+k}^{\mu\dagger} a_{r+k}^\nu\}.
\end{aligned}$$

When we substitute the values 6·4 for the components $A_{\mu\nu}$, $B_{\mu\nu}$ we see that only those terms survive for which $\mu = \nu$ and thus the $a_r^{\mu\dagger} a_r^\nu$ reduces to n_r^μ. In fact we have

$$\begin{aligned}
\xi_k \eta_k^\dagger - \eta_k^\dagger \xi_k = -i \sum_r \{&u(r, r+k)\, \bar{u}(r, r+k)\, (n_r^1 - n_{r+k}^2) \\
&+ u(r+k, r)\, \bar{u}(r+k, r)\, (n_{r+k}^1 - n_r^2)\}.
\end{aligned}$$

This can only vanish if the coefficient of each n_s^1, n_s^2 vanishes separately. I.e. if

$$|\, u(s, s+k)\, |^2 + |\, u(s, s-k)\, |^2 = 0,$$
$$|\, u(s+k, s)\, |^2 + |\, u(s-k, s)\, |^2 = 0.$$

The quantities occurring in these equations are essentially non-negative; each must therefore vanish separately:

$$u(r, r+k) = u(r+k, r) = 0.$$

This gives $$\xi_k = \eta_k = 0;$$

this, however, is inconsistent with (5·2). *The conditions can therefore not be satisfied*

Because of its generality, it is a little difficult to see through the foregoing argument, and it is of interest to start from a more special form for ξ_k, η_k in order to illustrate the nature of the failure. In the one-dimensional model, where ξ_k, η_k are replaced by one amplitude b_k and a_r^1, a_r^2 by α_r, the relation between b_k and α_r is (Born and Nagendra Nath 1936*a*, equation (37))

$$b_k = 1/\sqrt{k} \sum_{r=-\infty}^{\infty} \alpha_r^\dagger \alpha_{r+k}.$$

From this it is plausible to suppose that the $A_{\mu\nu}(r, r+k)$, $B_{\mu\nu}(r, r+k)$ are in fact independent of r and contain k only as $1/\sqrt{k}$.

$$\xi_k = 1/\sqrt{k}\, A_{\mu\nu} \sum_{r=-\infty}^{\infty} a_r^{\mu\dagger} a_{r+k}^{\nu},$$

$$\eta_k = 1/\sqrt{k}\, B_{\mu\nu} \sum_{r=-\infty}^{\infty} a_r^{\mu\dagger} a_{r+k}^{\nu}.$$

We shall see that by suitably choosing the matrices **A**, **B** we can satisfy the commutation rules (5·2), but that then it is impossible to satisfy (6·3).

As an example we calculate $\xi_k \eta_l^\dagger - \eta_l^\dagger \xi_k$. Let us write $B_{\sigma\tau}^\dagger = \bar{B}_{\tau\sigma}$; then

$$\begin{aligned}\sqrt{(kl)}\,(\xi_k \eta_l^\dagger - \eta_l^\dagger \xi_k) &= (A_{\mu\nu} \sum_r a_r^{\mu\dagger} a_{r+k}^{\nu})(B_{\sigma\tau}^\dagger \sum_s a_{s+l}^{\sigma\dagger} a_s^{\tau}) \\ &\qquad - (B_{\sigma\tau}^\dagger \sum_s a_{s+l}^{\sigma\dagger} a_s^{\tau})(A_{\mu\nu} \sum_r a_r^{\mu\dagger} a_{r+k}^{\nu}) \\ &= A_{\mu\nu} B_{\sigma\tau}^\dagger \sum_r \sum_s \{a_r^{\mu\dagger} a_{r+k}^{\nu} a_{s+l}^{\sigma\dagger} a_s^{\tau} - a_{s+l}^{\sigma\dagger} a_s^{\tau} a_r^{\mu\dagger} a_{r+k}^{\nu}\} \\ &= A_{\mu\nu} B_{\sigma\tau}^\dagger \sum_r \sum_s \{a_r^{\mu\dagger}(\delta_{\nu\sigma}\delta_{r+k,\,s+l} - a_{s+l}^{\sigma\dagger} a_{r+k}^{\nu}) a_s^{\tau} \\ &\qquad - a_{s+l}^{\sigma\dagger}(\delta_{\mu\tau}\delta_{rs} - a_r^{\mu\dagger} a_s^{\tau}) a_{r+k}^{\nu}\}.\end{aligned}$$

Those terms which are products of four amplitudes cancel out, and owing to the δ_{rs}, $\delta_{r+k,\,s+l}$ we can replace the double sum by a single sum:

$$\begin{aligned}&= A_{\mu\nu} B_{\sigma\tau}^\dagger \sum_r (a_r^{\mu\dagger} a_{r+k-l}^{\tau} \delta_{\nu\sigma} - a_{r+l}^{\sigma\dagger} a_{r+k}^{\nu} \delta_{\mu\tau}) \\ &= \sum_r (a_r^{\mu\dagger} A_{\mu\nu} B_{\nu\tau}^\dagger a_{r+k-l}^{\tau} - a_{r+l}^{\sigma\dagger} B_{\sigma\mu}^\dagger A_{\mu\nu} a_{r+k}^{\nu}).\end{aligned} \tag{6·5}$$

If $k \neq l$ we can split this up into the difference of two series:

$$= \sum_r a_r^{\mu\dagger} A_{\mu\nu} B_{\nu\tau}^\dagger a_{r+k-l}^\tau - \sum_r a_{r+l}^{\sigma\dagger} B_{\sigma\mu}^\dagger A_{\mu\nu} a_{r+k}^\nu, \tag{6·6}$$

for these two series converge. This can be rewritten as

$$\sqrt{(kl)}\,(\xi_k \eta_l^\dagger - \eta_l^\dagger \xi_k) = \sum_r a_r^{\mu\dagger} (AB^\dagger)_{\mu\nu} a_{r+k-l}^\nu - \sum_r a_r^{\mu\dagger} (B^\dagger A)_{\mu\nu} a_{r+k-l}^\nu.$$

A necessary and sufficient condition for this to vanish is

$$\mathsf{A}\mathsf{B}^\dagger = \mathsf{B}^\dagger \mathsf{A}. \tag{6·7}$$

If $k = l$ we cannot split the series into two parts, for the series in (6·6) diverge, but we can use the result (4·8).

$$\xi_k \eta_k^\dagger - \eta_k^\dagger \xi_k = T_r(\mathsf{A}\mathsf{B}^\dagger).$$

Hence $T_r(\mathsf{A}\mathsf{B}^\dagger) = 0$.

A similar investigation for the remaining products leads to the result that A, $\mathsf{A}^\dagger$, B, $\mathsf{B}^\dagger$ must all commute, and

$$\begin{aligned} T_r(\mathsf{A}\mathsf{A}^\dagger) &= T_r(\mathsf{B}\mathsf{B}^\dagger) = 1, \\ T_r(\mathsf{A}\mathsf{B}^\dagger) &= 0. \end{aligned} \tag{6·8}$$

Two examples of matrices satisfying these conditions may be quoted First, those studied by Nagendra Nath (1936, equation (16))*

$$\mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \quad \mathsf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0, & 1 \\ 1, & 0 \end{pmatrix}.$$

Second, those which occur in Kronig's theory (1936, equation (37))*

$$\mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1, & 0 \\ 0, & 1 \end{pmatrix}, \quad \mathsf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}.$$

A simple algebraic argument shows that matrices satisfying (6·8) cannot satisfy the invariance condition (6·3):

$$\begin{aligned} 2i\mathsf{B} &= \mathsf{A}\gamma - \gamma\mathsf{A}, \\ -2i\mathsf{A} &= \mathsf{B}\gamma - \gamma\mathsf{B}. \end{aligned}$$

For eliminating B from these equations

$$\begin{aligned} 4\mathsf{A} &= \gamma^2\mathsf{A} - 2\gamma\mathsf{A}\gamma + \mathsf{A}\gamma^2 \\ &= 2\mathsf{A} - 2\gamma\mathsf{A}\gamma, \\ \mathsf{A} &+ \gamma\mathsf{A}\gamma = 0. \end{aligned}$$

---

* Their notation is different; when put into the present notation their results reduce to this.

Multiplying on the left by γ

$$\gamma A + A\gamma = 0.$$

Since γ is Hermitian we also have, on taking the adjoint of this,

$$\gamma A^\dagger + A^\dagger \gamma = 0.$$

From the first of the equations (6·3)

$$B = i\gamma A.$$

But by (6·8), $A^\dagger$ and B commute with one another, hence

$$\begin{aligned} 0 &= BA^\dagger - A^\dagger B \\ &= i(\gamma AA^\dagger - A^\dagger \gamma A) \\ &= i(\gamma AA^\dagger + \gamma A^\dagger A). \end{aligned}$$

Multiplying by $-i\gamma$ $\qquad AA^\dagger + A^\dagger A = 0,$

i.e. $\qquad T_r(AA^\dagger) = T_r(A^\dagger A) = 0,$

which contradicts (6·8).

I am indebted to Professors J. v. Neumann, P. Jordan and R. de L. Kronig and to Mr N. S. Nagendra Nath for helpful discussion.

SUMMARY

This paper brings to light a grave difficulty for the neutrino theory of light. Starting from assumptions about the neutrino sufficiently general to include the models which have been studied by Jordan, Kronig and others (with the exception of Scherzer's attempt, which is not strictly a neutrino theory), and working with the amplitudes of the second quantization as the most suitable mathematical apparatus, one sets up the most general theory consistent with Jordan's hypothesis. The conditions under which this will lead to a satisfactory theory of light are (1) that certain commutation rules be satisfied, (2) that the theory be invariant under a change of co-ordinate system. In order to study the second of these it has been necessary to analyse rather carefully the transformation of the amplitudes under certain types of rotation and this reveals an arbitrariness in the choice of certain phases. A condition for the invariance of the theory is that the results be independent of the way in which these phases are chosen.

From this point onward a straightforward analysis leads to the result that the conditions cannot be satisfied simultaneously. The invariance requires that the neutrino which interacts with the atom should reverse its spin, a result which could also be derived from considerations of the conservation of angular momentum, and an essentially simple though rather tedious calculation shows this to be inconsistent with the commutation rules.

The introduction gives an account of the aims of the neutrino theory of light, the problems which it meets and the attempts that have been made to solve them.

REFERENCES

Born and Nagendra Nath 1936 *a* *Proc. Ind. Acad. Sci.* **3**, 318
— — 1936 *b* *Proc. Ind. Acad. Sci.* **4**, 611.
de Broglie 1932 *C.R. Acad. Sci., Paris*, **195**, 862, **197**, 536
— — 1933 *C.R. Acad. Sci., Paris*, **197**, 1377.
— — 1934 *C.R. Acad. Sci., Paris*, **199**, 445; **199**, 1165
de Broglie and Winter 1934 *C.R. Acad. Sci., Paris*, **199**, 813.
Dirac 1930 "Principles of Quantum Mechanics." Oxford. Clarendon Press.
Fock 1937 *a* *Phys. Z. Sowjet* **11**, 1.
— 1937 *b* *C.R. Acad. Sci. U.R.S.S.* **4**, 229.
Jordan 1935 *Z. Phys.* **93**, 464
— 1936 *a* *Z. Phys.* **98**, 759.
— 1936 *b* *Z. Phys.* **99**, 109.
— 1936 *c* *Z. Phys.* **102**, 243.
— 1937 *a* *Z. Phys.* **105**, 114.
— 1937 *b* *Z. Phys.* **105**, 229
Jordan and Kronig 1936 *Z. Phys.* **100**, 569.
Jordan and Wigner 1928 *Z. Phys.* **47**, 631
Kronig 1935 *a* *Physica*, **2**, 491
— 1935 *b* *Physica*, **2**, 854.
— 1935 *c* *Physica*, **2**, 968.
— 1936 *Physica*, **3**, 1120
Nagendra Nath 1936 *Proc. Ind. Acad. Sci.* **3**, 448.
Scherzer 1935 *Z. Phys.* **97**, 725.
Sokolow 1937 *Phys. Z. Sowjet.* **12**, 148

The absorption spectra of sulphur dioxide and carbon disulphide in the vacuum ultra-violet

By W. C. Price and (Miss) D. M. Simpson

Physical Chemistry Laboratory, Cambridge

(*Communicated by R. G. W. Norrish, F.R.S.—Received 23 December 1937. Revised in proof 26 February 1938*)

[Plates 3, 4]

In the near ultra-violet the absorption spectrum of sulphur dioxide has been investigated to a greater extent than that of any other triatomic molecule. The excellent work of Clements (1935) on the temperature dependence of the bands has enabled the ν_0 of the upper state to be definitely fixed. It has resulted in a satisfactory analysis of the so-called low-frequency system (i.e. bands appearing at high pressures which are due to transitions from various initial vibrating states), and has yielded a plausible arrangement of part of the high-frequency system (bands appearing at low pressures and temperatures, and probably corresponding to transitions from vibrationless ground states to the various vibrational levels of the upper state). Asundi and Samuel (1935) have put forward an alternative analysis of these bands, but we do not favour it on the grounds that it disregards the results of temperature experiments, contains many violations of Herzberg and Teller's selection rules (1933), and interprets several strong bands as transitions from initial vibrational states in spite of prohibitive Boltzmann factors. A photograph of the spectrum is shown in fig. 1*b*, Plate 3. The bands are very strong and appear at pressures of about 1/2 mm. in a path length of 1 m. Another system several times stronger than the previous one starts in the region of 2350 A. It has been investigated by several experimenters, but the only attempt at an analysis has been made by Chow (1933*a*, *b*). One of the difficulties that has troubled previous experimenters is that the bands continue to shorter wave-lengths past the transmission limit of quartz, and thus go outside the range of their instruments. We have therefore photographed the bands with a vacuum grating spectrograph, and in this way have obtained the absorption spectrum of sulphur dioxide down to about 1000 A. The technique used in obtaining the absorption spectra has been described previously (Collins and Price 1934). The Lyman continuum was employed as the continuous background against which the absorption was observed.

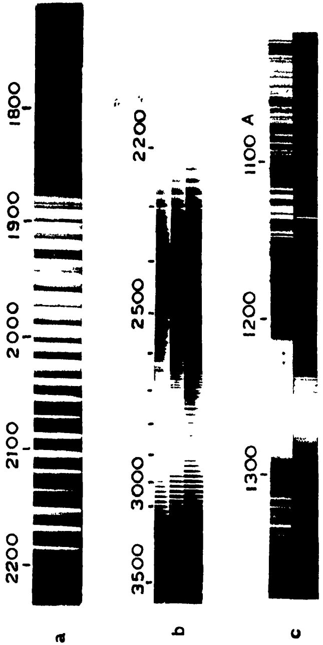

1800
1900
2000
2100
2200
2200
2500
3000
3500
1100 A
1200
1300
a
b
c

FIG. 1

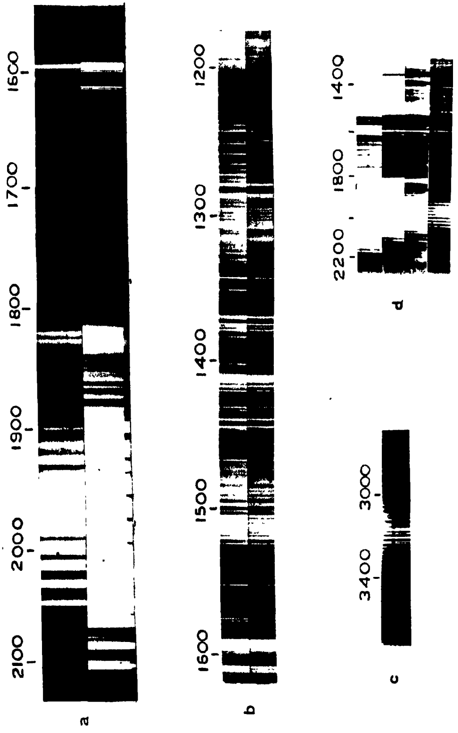

1600
1700
1800
1900
2000
2100
a
1200
1300
1400
1500
1600
b
3000
3400
c
1400
1800
2200
d

FIG. 3

A photograph of the 2350–1800 A system of bands is given in fig. 1*a*, Plate 3. They are shaded towards the red, and many of them possess more than a single head. As far down as 1950 A some degree of regularity is apparent in the spectrum.† Beyond this the bands depart from a regular spacing, and tend to split up into several components in a rather peculiar way. Between 2300 and 1950 A we find about twenty bands forming a progression with an average spacing of 380 cm.$^{-1}$ (see Table I). The separation between consecutive members, which is taken between the centres of gravity of the various bands, is found to remain constant within quite narrow limits. After trying to fit the bands into many alternative schemes it has been concluded that a simple progression with a 380 cm.$^{-1}$ difference accounts most satisfactorily for all the strong bands which appear in this region. The separation 380 cm. ^{1} most probably corresponds to the deformation frequency ν_2, which has a value of ~524 cm $^{-1}$ in the ground state. It shows little tendency to fall off to a smaller value in going to shorter wave-lengths, and as the work of Clements (1935), Lotmar (1933), and Chow (1933*a*, *b*) indicates that the anharmonic constant for the 524 cm.$^{-1}$ frequency is extremely small, the similar behaviour shown by the 380 difference would seem to lend support to the suggestion that it also corresponds to the ν_2 mode of vibration. On account of the irregularities and the extremely small value of the anharmonic constant no attempt has been made to express the bands by a formula Temperature variation experiments, which will be published separately, indicate that the 43,100 cm.$^{-1}$ band is the first member of the progression.

There are two main difficulties which the above interpretation of the bands encounters. The first is that there is a periodic fluctuation in the intensity of the bands. Starting with the band at 2206 A, and proceeding to shorter wave-lengths, the intensity of the main members gradually increases up to the fourth and then drops suddenly to the fifth. Subsequently it rises again to the eighth, and drops at the ninth. There is a further rise to the twelfth, and drop at the thirteenth, and this is repeated at the sixteenth and seventeenth bands respectively Now all these strong bands almost certainly originate on the vibrationless ground state. If the bands form a single progression, it might be expected from analogy with the diatomic case, using the Franck-Condon principle, that the intensity of the bands would increase uniformly to a maximum, and die away on either side of this. However, if the transition probabilities are worked out by wave mechanical methods, it can be predicted that subsidiary maxima may accompany the main one (Hutchisson 1930, 1931). This is found to occur for instance in the Dieke and Hopfield bands of hydrogen. The peculiar intensity fluctuations may therefore have some explanation on such a basis. Other more likely causes will be discussed later.

---

† We do not find any very definite evidence for the sudden onset of predissociation at this wave-length as reported by Henri (1931). Certainly there is some general diffuseness, but the increase in multiplicity and the departure from regularity are the most striking features of these shorter wave-length bands.

TABLE I. FREQUENCIES IN THE 380 CM.$^{-1}$ PROGRESSION OF THE 2300–1800 A SYSTEM IN SO_2

I	ν cm.$^{-1}$	$\Delta\nu$ cm.$^{-1}$	I	ν cm^{-1}	$\Delta\nu$ cm.$^{-1}$
0	43,100	—	7	47,220	380
1	43,470	370	9	47,590 *M*	370
1	43,840	370	9	47,980 *M*	390
1	44,220	380	7	48,355	375
2	44,600	380	9	48,730	375
3	44,980	380	9	49,120	390
5	45,350 *M*	370	10	49,500 *M*	380
7	45,740	390	10	49,890	390
8	46,110	370	10	50,280	390
8	46,470	360	10	50,670	390
5	46,840 *M*	370	10	51,030	360

$M \equiv$ mean values

The second difficulty is that the internal structure of the individual bands varies considerably This gives rise to some apparent irregularities in the measured spacing. Some of the bands are double, while others seem to possess only a single head though the structure is not in general so simple as this remark would imply. The sulphur dioxide molecule is an asymmetric top rotator, and the bands belonging to it should therefore be expected to have a complicated rotational structure which will depend to some extent on the vibrational state of the molecule. We have found it misleading to compare the structures of two bands which appear on the same plate with very different intensities. In fact, if these bands are examined on two plates, which are taken at such pressures that the weaker band on the high-pressure plate appears at the same strength as the stronger band on the low-pressure plate, then it is generally found that the difference in structure is not nearly so marked.

Chow in his analysis has resorted to the device of splitting up the system into three neighbouring electronic states in order to arrange the bands. However, we consider it desirable, for reasons which will be apparent later when the bands are compared to those at longer wave-lengths and to similar systems in carbon disulphide, to interpret them as a single system.

Minor irregularities in the separations and structures are explained as due to the influence of neighbouring states. Clements (1935) uses such an explanation to account for the deviations which occur in his ν_2 (220 cm.$^{-1}$) progression. In referring to the disturbance in the neighbourhood of the band "O", he states that "perhaps this is due to an interaction of the normal modes of vibration in the upper state, since ν_1 and ν_2 and even multiples of ν_3 belong to the same representation of the symmetry group". With increasing amplitudes of the molecule corresponding to the higher vibrational quantum numbers, the interaction may be supposed to increase, so that this explanation may be called upon to account for the much greater irregularities which occur around 2600 A. The breadth and diffuseness of the bands in the region 2800–2500 A has been tentatively attributed by Franck, Sponer and Teller (1932) to collision broadening, they pointed out that energy considerations prevented it from being due to predissociation. The chief difficulty with their explanation is that it requires the molecule in the excited state to have a collision radius 15 times greater than in its normal state.

The short wave-length bands of the 2300–1800 A system behave in a rather similar way to the corresponding members of the 3300–2600 A system. It would clearly be advantageous to have a common explanation for both, especially as the corresponding bands in carbon disulphide also seem to exhibit the same kind of peculiarity. It has been remarked by Mulliken that there must be considerable interaction between the various modes of vibration of a polyatomic molecule, and that the variables referring to the normal modes cannot in general be completely separable. Effects resulting from this will become increasingly important with the increase in the number of vibrational quanta present. Any particular vibrational band A of an electronic state in a polyatomic molecule is liable to be perturbed by a state B provided the symmetries of the vibronic wave functions of both states belong to the same species (Mulliken 1937). B would, of course, be built up from a different combination of the proper frequencies of the electronic state and the magnitude of the perturbation would depend on the proximity of the energy values of A and B. Such a perturbation could be used to account for the variation in intensity observed among the lower members. For example, let us consider a polyatomic molecule which has a low vibrational frequency (e.g. 300 cm.$^{-1}$) the others being high (1000 cm.$^{-1}$) and for simplicity further assume that both 300 and 1000 cm.$^{-1}$ frequencies are totally symmetrical. In a long progression of the 300 frequency, one might expect the tenth member to be perturbed by the third member of the 1000 frequency, provided that the anharmonic constants were not too large. If the higher frequency happened to be roughly a multiple of the lower one, in this case for example 900 or 1200 cm.$^{-1}$, then a periodic perturbation would be apparent in the progression of the smaller vibration. Thus the variation in intensity with every fourth member in the 2300–1800 A system may be due to a perturbing frequency in the upper state of about 1500 cm.$^{-1}$, which belongs to the same vibrational species as ν_2, e.g. probably ν_1. Similarly, it seems possible that the 1320 cm.$^{-1}$ upper state frequency of Clements (1935) arises from a regular perturbation frequency of this magnitude rather than from one which is really present in the spectrum. If this is not so, it is necessary to assume that the 1320 is an exact multiple of the 220 cm.$^{-1}$ frequency, in the same way that it would be necessary for one frequency to be four times the other to account for the periodicity of the 2300–1800 A system. The much more violent disturbances which occur below 1950 A in the latter system may result from the fact that for such high vibrational states more combinations of the normal frequencies can have energy values laying close to the unperturbed state, and are thus able to affect it. Such an explanation might well account for the complexities of the short wave-length bands of both the 3300–2600 A and the 2300–1850 A systems.

The assignment of both the 220 cm.$^{-1}$ difference in the 3300–2600 A system and the 380 cm $^{-1}$ difference in the 2300–1850 A system to the angle vibration of sulphur dioxide, can be supported to some extent by the consideration of a simplified potential energy diagram. Possible U/θ curves for the ground, first excited and second excited states of sulphur dioxide are shown schematically in fig. 2, θ being the $\widehat{OSO}$ angle. To obtain some idea of the absorption spectrum arising from transitions between the lowest level of the ground state vertical lines a, b, c, and α, β, γ are drawn from $a\alpha$ (the zero point energy $\frac{1}{2}\nu_2$) cutting the upper potential curves in $b\beta$ and $c\gamma$. The vertical distances between b and β, and c and γ give a measure of the frequency range over which the absorption spectrum is likely to extend. It can be seen from the figure that a much greater frequency spread is to be expected for the state in which the vibration frequency is little diminished, than for that in which it is considerably reduced.† The 3300–2600 A system in which ν_2 has fallen to the low value of 220 cm.$^{-1}$ has a frequency range of only 6000 cm.$^{-1}$; on the other hand, the 2300–1850 A system extends over the much greater spread of 13,000 cm.$^{-1}$. It therefore appears that the correlation of the 220 and

---

† The above argument is only valid when the potential curves of both the upper states have changed considerably from that of the ground state. This seems certainly to be true in the case under discussion.

380 cm.$^{-1}$ frequencies is supported in some measure by the frequency range over which their respective states extend.

The probable explanation of the origin of the bands seems to be that they are due to the transitions of a relatively non-bonding electron localised on an oxygen atom (i.e. the "lone pair" $2p_0$) into excited orbitals which are anti-bonding in that they weaken the angular restoring forces.† Since the excited orbital of the longer wave-length system is not so attenuated as that of the shorter wave-length system, but is closer in and concentrated between the atoms, its anti-bonding power is likely to be the larger one, if all the other factors are the same. A similar explanation seems to account for the breadths of the analogous systems in carbon disulphide. Mulliken's discussion of the electronic structure of sulphur dioxide (Mulliken 1935c) shows that the most loosely bound electrons are slightly anti-bonding and would, in fact, be strongly so were it not for the triangular configuration of the molecule. The preference is given to the non-bonding electrons as originators of the absorption, because of the apparent absence of strongly excited valence frequencies in the upper state. The alternative explanation that the excitation is that of a bonding (valence) electron to an anti-bonding orbital, after the manner of the Schumann-Runge bands of oxygen, would probably have to be applied if the molecule happened to be linear.

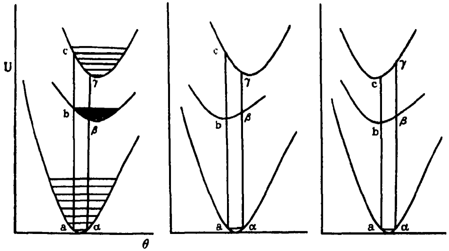


Fig. 2. Schematic U/θ curves for SO_2.

The $\widehat{OSO}$ angle as determined in the ground state from electron diffraction measurements is 124 ± 15° (Brockway 1936). Jonescu (1933) from a partial rotational analysis reports the $\widehat{OSO}$ angle in the first excited state to be 96°. This computation assumes that the SO distances are not greatly altered by the excitation. Such an assumption could not in general be valid. However, according to the proposed analysis the so-called valence vibrations do not appear strongly in the spectrum This suggests that the valence fields are not much changed by the excitation, and Jonescu's assumptions might be expected to hold roughly. Certainly the fact that he is able to deduce the $\widehat{OSO}$ angle in the ground state to be 120° supports them to some extent. It can be further tested by the following rough calculation. If we take the number of ν_2 quanta contained in the range between the ν_0 and the intensity maximum of the system it is possible to calculate an approximate value of the change in the angle. On the Franck-Condon principle the intensity maximum of the absorption bands corresponds to a transition to a vibrating upper state in which the greatest, or least angle attained is equal to that in the non-vibrating ground state Assuming (I) a parabolic U/θ curve, (II) that the SO distances are not greatly changed by the excitation, (III) that the angle vibration in the upper state has a frequency of 220 cm $^{-1}$, (IV) that the distance between ν_0 (31,945) and the intensity maximum is 3300 cm $^{-1}$ (i e about $15\nu_2$), the change in the angle works out to be approximately 28°, a result which is in good agreement with that suggested by Jonescu (24°). Using similar assumptions for the 2000 A system the change in angle works out to be about 22°.

---

† This does not necessarily mean that θ_e increases.

One of the most definite characteristics of both the 3000 and 2000 A systems of sulphur dioxide is that they are shaded towards the red. For molecules of the asymmetric top type, it is not in general possible to draw any conclusion concerning the change of a particular moment of inertia from this fact, without a more detailed analysis of the structure, such as has presumably been carried out by Jonescu. However, since the moments of inertia of sulphur dioxide in the normal state are roughly 12, 73, and 85×10^{-40} g.cm ^{2}, it can be regarded approximately as a symmetrical top rotator, in which one moment of inertia is very considerably less than the other two. The gross structure in this case would be expected to consist of a number of subheads. The subheads themselves would be shaded in the direction suggested by the change of the large moments of inertia, while the shading of the gross structure would refer to the change in the small moment of inertia. Inspection of the sulphur dioxide bands (both of the 3000 and the 2000 A systems) indicates the presence of subheads, which are frequently very narrow and for which the direction of shading cannot usually be specified with certainty. The gross structure is, however, definitely shaded towards the red. Thus it would appear that the small moment of inertia is increased by the excitation. This is most probably explained as a diminution in the $\widehat{OSO}$ angle, in accordance with the calculations of Jonescu.

On the short wave-length side of the 2300–1850 A bands there is a region of transparency. Below 1600 A we come to the next electronic system. This consists of three rather weak diffuse bands, whose approximate wave-lengths are 1573, 1558, and 1529 A. It is difficult to know whether to class these bands with those at longer wave-lengths as having their upper states represented by anti-bonding orbitals, or to group them with the bands at still shorter wave-lengths, which are of a Rydberg type. The former interpretation is considered to be the most likely one. It should also be mentioned that in the region 2700–2400 A, there seems to be evidence of a separate electronic transition. If this is so, then it would appear that sulphur dioxide has at least four electronic transitions above 1500 A.

The bands in sulphur dioxide, which might be regarded as resonance bands approaching the ionization potential, occur below 1350 A They have a complicated appearance and comparatively little can be done with the component bands of the individual electronic states This complexity may possibly result from strong resonance between OSO* and O*SO, which is likely to disturb some of the higher excited states. In spite of this, it is fairly easy to pick out the regions of absorption corresponding to separate electronic transitions For instance, the bands between 1350 and 1308 A obviously correspond to a different electronic transition from those between 1280 and 1240 A, both on account of their appearance, and of the well-defined break in the absorption The comparatively narrow frequency spread of the electronic states indicates that an electron of not very large bonding, or anti-bonding power is being removed Between 1215 and 1140 A there is a wide region of transparency, after which there appears a set of weaker bands. It is almost certain that these bands are the successive Rydberg members of a series starting with the absorption in the 1350–1215 A region, especially as still weaker bands corresponding to the higher members of such a set appear at shorter wave-lengths. A rough Rydberg extrapolation indicates that the ionization of the molecule will occur at 12·05 ± 0·05 V. This is supported by the fact that the regions of absorption of the bands of sulphur dioxide occur a little to the long wave-length side of the corresponding ones in water.† (Absorption also assumed to arise from transitions between orbitals, which are closely atomic (non-bonding) in character.) The extrapolation of the Rydberg series in water leads to an ionization potential of 12·56 V (Price 1936*a*). A value of 12·05 V for the ionization potential of sulphur dioxide is therefore in agreement with its spectrum. The electron impact value of 13·1 ± 0·3 V (Smyth and Mueller 1933) seems a little high. There is no doubt that this ionization potential corresponds to the removal of an electron which has little bonding power, and which is roughly a $2p_0$ electron The predicted value for such an electron is 14·7 V (Mulliken 1934). Part of the discrepancy is to be attributed to the $\overset{-}{\mathrm{O}}\overset{+}{\mathrm{S}}\overset{-}{\mathrm{O}}$ polarity (dipole moment 1·6 *D*), the accumulation of negative charge in the oxygen atoms reducing the ionization potentials of the electrons localized on them. Part of it, however, is probably a result of the electron retaining a certain amount of the anti-bonding power, which it would possess in a very marked degree, if the molecule were linear (Mulliken 1935*c*). The electron configuration of molecules having sixteen outer electrons (e g. CO_2, CS_2 and N_2O in their normal states) consists in a set of firmly bound closed shells, giving a chemically saturated structure like that of N_2. Additional electrons introduced into such a structure, in an attempt to get the configuration of NO_2 (one more) or SO_2, Cl_2O, O_3 (two more), would have to go into anti-bonding orbitals, as in the somewhat similar diatomic cases of NO and O_2 However, as Mulliken has pointed out, the triatomic molecules given above avoid this by assuming a triangular form, and in this way acquire a different set of orbitals. In the triangular configuration, it seems that the additional electrons have relatively little anti-bonding power, since, at least for SO_2, the diminution in the ionization potential is not very much greater than that which might be expected from charge transfer effects alone.

---

† Similarities can be traced between the AB series in water, and the observed bands in sulphur dioxide.

THE ABSORPTION SPECTRUM OF CARBON DISULPHIDE

The absorption spectrum of carbon disulphide, though it differs in many of its fundamental characteristics from that of sulphur dioxide, has nevertheless several important features in common with it. These are particularly prominent for the bands lying at wave-lengths greater than 1800 A. The first moderately intense electronic absorption of carbon disulphide occurs at 3600–2900 A, see fig. 3*c*, Plate 4. It appears strongly only when pressures greater than several centimetres are used, the path length being ~1 m. This is much weaker than the corresponding system in sulphur dioxide (probably by a factor of 50). Mulliken (1935*c*) has tentatively assigned this system to a certain forbidden transition, stating that it becomes allowed because of the possibility of the molecule assuming a triangular form in the excited state. The selection rules for the symmetry $D_{\infty d}$ are then replaced by those of C_{2v}, and the transition becomes possible as in the case of sulphur dioxide. The appearance of the spectrum is more complicated than that of sulphur dioxide, and its analysis has not yet been satisfactorily accomplished.

A much more intense system occurs at shorter wave-lengths, see fig. 3*a*, Plate 4. In the region 2200–1800 A there is a considerable degree of regularity apparent among the bands. Inspection of the small photograph (Fig. 3*d*) indicates the presence of a many membered vibrational progression, which rapidly widens with increase in pressure. (These photographs were taken on a small exploratory spectrograph, and serve mainly to give a bird's eye view of the spectrum, and to show its variation with pressure.) It can be seen from the high dispersion picture that each band is really built up from many components, the band at 2080 A, for example, shows several heads Their narrowness is in part accounted for by the very large moment of inertia of carbon disulphide. Of all the possible explanations for the multiplicity of these bands, we consider that the best is that they are due to a rotational structure, which arises as a result of the molecule becoming very slightly bent in the upper state, and thus acquiring one relatively low moment of inertia It may be mentioned here that the separation of the heads is about 40 cm $^{-1}$ and that the structure as a whole appears to be shaded towards the red This is compatible with the hypothesis that the molecule is bent through an angle of two or three degrees as a result of the excitation A similar explanation will be given later to account for the structure of the band at 1600 A.

The advance of the 2200–1850 A progression towards longer wave-lengths with increase in pressure is at first quite rapid, and it reaches 2150 A with pressures of 0·5 mm. and a 1 m. path length. To extend the absorption further requires much higher pressures. At 40 cm the bands can be followed as far as 2300 A, and the system shows signs of continuing still further As in the case of sulphur dioxide, it was found that when the centres of gravity of the bands are taken, the gross structure forms a fairly good progression with only slight variations in the frequency separations of consecutive members Their frequencies are given in Table II, the 45,950 cm.$^{-1}$ band is indicated as the first member by temperature variation experiments. It is even more certain in this case, than in that of sulphur dioxide, that we are dealing with a single electronic upper state, and with a vibrational progression involving only one type of vibration. The frequency difference is found to be about 410 cm.$^{-1}$ (varying between 390 and 430 cm.$^{-1}$). This is undoubtedly to be assigned to the symmetrical valence frequency ν_1, which has a value of 656·5 cm.$^{-1}$ in the ground state. The reasons for this assignment are as follows: the appearance of the spectrum indicates that, as a result of the excitation, there has been a change in the vibration frequency; this is almost certainly a diminution. From the fact that the bands appear at such low pressures, it is certain that the electronic transition is an allowed one. The selection rules then permit only the totally symmetrical vibration frequencies to appear in absorption from the vibrationless ground state. As the only totally symmetrical frequency in carbon disulphide is ν_1, both the diminution in the frequency, and the selection rules point to this assignment for our 410 cm.$^{-1}$ separation. The absence of periodic fluctuations in the intensity of the bands is probably to be connected with the fact that neither of the other two vibrations of the molecule has the correct symmetry for causing such a perturbation. Some sixteen bands are included in the progression, which extends over a range of more than 10,000 cm.$^{-1}$. This range is considerably greater than that of the near ultra-violet system, which at pressures of several centimetres covers only 6000–7000 cm.$^{-1}$. Great increase in pressure will extend the latter system by about 1000 cm.$^{-1}$, but this is, of course, largely a result of transitions from the vibrating ground states. The same arguments apply here as were put forward for sulphur dioxide, to suggest that the "3000 A" bands correspond to a greater anti-bonding effect and to lower vibrating frequencies than the "2000 A" bands. Since all authors who have attempted to analyse the near ultra-violet system have found evidence of a frequency of about 250 cm.$^{-1}$, it seems possible that this corresponds to ν_1 in the excited state (Wilson 1929, Watson and Parker 1931 *a*, *b*; Jenkins 1929). However, from the complicated appearance of the spectrum, it is certain that more than one upper state frequency is strongly excited, and therefore no definite statement can be made until a complete analysis of the spectrum is forthcoming. It should be pointed out that if the molecule is bent in this excited state, as appears likely from electronic selection rules (Mulliken 1935 *c*), and also from the probable appearance of more than one upper state frequency, then the angle through which it is bent is fairly large (~10°), as the narrowness of the bands indicates a high moment of inertia. However, this argument would not hold if the change in electric moment were along the SCS axis. The wide rotational structure of the 2000 A, 1600 A and possibly other systems indicates that the change in electric moment involved in these transitions is perpendicular to the SCS axis.

Proceeding to shorter wave-lengths, it is most noticeable that the strong band at 1815 A differs in type from the preceding ones. Since it is obviously unaccompanied by any vibrational structure, it must be assigned to a transition between non-bonding atomic orbitals (possibly $3p \rightarrow 4s$). Though somewhat diffuse, it resolves itself at low pressures into two components (λ1818·1 and 1811·7 A). A similar band occurs at 1600 A where the non-bonding type is even more striking.

TABLE II. FREQUENCIES IN THE 410 CM $^{-1}$ PROGRESSION OF THE 2300–1800 A SYSTEM IN CS_2

I	ν cm.$^{-1}$	$\Delta\nu$ cm.$^{-1}$	*I*	ν cm.$^{-1}$	$\Delta\nu$ cm.$^{-1}$	
00	45,950	—	6	49,290	410	
0	46,370	420	7	49,710	430	
0	46,790	420	7	50,110 *M*	400	385
1	47,200	410	8	50,480	370	
1	47,610	410	9	50,900	420	
2	48,020 *M*	410	10	51,280	380	410
3	48,440 *M*	420	8	51,720	440	
5	48,870	430	7	42,140	420	

M ≡ mean values.

TABLE III. FREQUENCIES IN THE 830 CM.$^{-1}$ PROGRESSION OF THE 1750–1650 A SYSTEM IN CS_2

I	ν cm.$^{-1}$	$\Delta\nu$ cm.$^{-1}$
3	56,490	—
10 (2)	57,320	830
10 *D*	58,170	850
8 *DD*	59,000	830
2 *DD*	59,820	820
1 *DD*	60,640	820
0 *DD*	61,420	780

To the short wave-length side of the band at 1800 A appear three weak bands, see fig. 3*a*, Plate 4. These form a progression with a frequency separation ~830 cm.$^{-1}$ With considerable increases in pressure more members appear, and it was found possible to measure seven in all, on very high pressure plates. As it seems probable that the vibration frequency to which this separation corresponds has been diminished by the excitation, it might conceivably be the ν_3 type, which has a value of 1524 cm.$^{-1}$ in the ground state. The objection to this interpretation is that this particular vibration is not a totally symmetrical one. However, it is conceivable that it might be allowed to appear because of dissymetry caused by the excitation SCS to SCS*. This will be explained later, when a similar vibration associated with the band at 1600 A is discussed. The possibility that the bands may be due to CS formed by photodissociation, or even to small traces of COS impurity, cannot be entirely ruled out.

The strong band which occurs at 1600 A is probably the most interesting in the whole spectrum. Since only very weak bands occur in the range ± 2000 cm.$^{-1}$ on either side of this band, it must obviously be assigned to the excitation of a non-bonding electron. That this electron is extremely non-bonding (i.e. the transition approximates to one between atomic orbitals) is shown by the faintness of the accompanying vibration bands on the short wave-length side. The band towards longer wave-lengths arises as a transition from a vibrating ground state, and its intensity is therefore governed by the appropriate Boltzmann factor. The wave number separation of the centres of gravity of this band and the main one is 650 cm.$^{-1}$, which corresponds to a transition from the ν_1 (656·5 cm.$^{-1}$) vibration from the ground state. A transition from the lower frequency ν_2 (396·8 cm.$^{-1}$) is not observed, as it is forbidden by selection rules. A band $2\nu_2$ might be expected, but apparently the Boltzmann factor prevents its appearing. To the short wave-length side of the main band, and separated from it by 660 cm.$^{-1}$, there occurs another weak band. This separation must undoubtedly correspond to a vibration of the ν_1 in the upper states, an assignment which is in agreement both with the vibrational selection rules, and with the fact that the electronic excitation is that of a non-bonding electron (ν_1 changes only from 660 to 650 cm.$^{-1}$, i.e. 10 cm $^{-1}$, as a result of an electron jump of nearly 8 electron volts, or 63,000 cm.$^{-1}$). A still weaker band appears at 1577 A, separated from the main band by about 1670 cm.$^{-1}$. This probably corresponds to the frequency ν_3, 1523 cm.$^{-1}$ in the ground state. While ordinarily it would be a forbidden transition, it seems to appear as a result of the dissymmetry induced in the molecule by the excitation (SCS to SCS*). This makes the symmetries of ν_1 and ν_3 equivalent as for example in SCO, where the analogous vibrations appear both in Raman spectra and the infra-red.

One of the puzzling features of the bands of this group is their well-resolved structure. They are composed of a number of very narrow heads (some of which are doublets $\Delta\nu \sim 21$ cm.$^{-1}$) whose frequencies can be determined accurately to two or three wave numbers. There are two possible explanations for this structure: firstly it might be caused by transitions from initial vibrating states which retained their original type of vibrational quanta in the excited state. The slight modification in the value of the frequencies in the upper state will cause a displacement of the bands relative to the corresponding ones arising from a vibrationless ground state. Such transitions should be relatively weak, and cannot account for the fact that both the band at 1595 and that at 1577 A have two equally intense heads as their main components. Further, if the explanation is so simple, the two strongest members of the 1612 and 1595 A bands should be separated by exactly 656·5 cm.$^{-1}$, whereas no two components have been found with exactly this difference. Our frequency of 650 cm.$^{-1}$ was found by averaging over each band. The second, and probably the correct, explanation is that the structure is rotational in origin. If the molecule becomes very slightly bent in the upper state it can, in this way, acquire a sufficiently low moment of inertia to give a wide rotational spacing. The observed separations are ~70 cm.$^{-1}$, such as would occur if the carbon disulphide molecule deviated from its linear configuration by a degree or two, and so approximated to a symmetrical top rotator. The fact that the structure appears shaded towards the red would also support this theory. As others might desire to analyse further the bands in the 1600 A region, a table of frequencies correct to 2 or 3 cm^{-1} is given in Table IV.

TABLE IV. FREQUENCIES IN THE SYSTEM 1012–1550 A IN CARBON DISULPHIDE

1612 A band		1595 A band		1877 A band		1553 A band	
I	ν cm.$^{-1}$	*I*	ν cm.$^{-1}$	*I*	ν cm.$^{-1}$	*I*	ν cm.$^{-1}$
00	61,865	0	62,250	0	63,204	0	64,324
1	61,960	3	62,626	0	63,284	2	64,398
1	62,017	8	62,702	2	63,360	2	64,416
3	62,080	10	62,754	2	63,430	0	64,452
		10	62,774	00	63,448		
		3	62,814	0	63,512		
				0	63,350		

Between 1535 and 1450 A there occurs a complicated set of bands crowded together without any obvious regularities. It is felt certain that no conceivable vibrational pattern will explain the character of this set. If the carbon disulphide molecule is very slightly bent in the excited state by a fraction of a degree, then part of the structure observed may well be rotational. On the other hand, resonance between the equivalent states S*CS and SCS* may be responsible for the complexity. The dimensions of the excited orbital seem to be of the right order for producing this effect. Below 1450 A the bands are less complicated, and obviously correspond to a large number of different electronic transitions, with little or no accompanying vibrational structure. Since, with the possible exception of a few bands in the neighbourhood of 1250 A, the multiplicity of heads characteristic of the 1600 and 2000 A systems is not a feature of the shorter wave-length bands, it is most probable that the molecule is linear in these highly excited states, as it is in the ion. The bands are fairly strong, and well separated at long wave-lengths, but become weaker and crowd closer together towards shorter wave-lengths, after the manner of bands going to an ionization potential. They converge to a limit around 1230 A, or ~10 V. As the electron impact value of the ionization potential of carbon disulphide is 10·4 ± 0·2 V (Smyth and Blewett 1934), the above interpretation of the bands is highly probable. Starting from the short wave-length limit of the bands, it was found possible to pick out the members of the following Rydberg series:

$$\nu_0^n = 81{,}734 - R/(n+0{\cdot}55)^2 \quad (n = 3,\ 4,\ 5,\ \text{etc.}). \tag{1}$$

Bands were found to fit in the formula from $n = 3$ to $n = 14$. Their frequencies are given in Table V. In spite of the complexity of the bands near the limit it is felt that there can be little doubt as to the reality of this series. The limit corresponds to 10·083 V, and the possible error in its determination is considered to be less than 0·005 V. It corresponds to the removal of a $3p\pi_g$ (or $(\pi-\pi,\ \pi_g)$) electron from the $^1\Sigma_g^+$ state of carbon disulphide (Mulliken 1935*c*). This leaves the molecular ion in a $^2\Pi_g$ state The doublet separation in the analogous case of CO_2^+ is 161 cm.$^{-1}$. By a consideration of the relative values of the spin-orbit coupling coefficients of O, O^+ and S, S^+ and using the above value for carbon dioxide, it is possible to predict that the doublet separation for the $^2\Pi_g$ state of CS_2^+ should be about 400 cm.$^{-1}$. A search among the shorter wave-length bands revealed many differences of ~440 cm $^{-1}$. Thus it appears that the doublet separation of the ion shows up in the highly excited states of the molecule, in the same way as it does for the alkyl halides (Price 1936*b*). In fact, each member of series (1) was found to possess a companion situated ~440 cm.$^{-1}$ on its long wave-length side These latter bands are fairly well represented by the following formula:

$$\nu_0^n = 81{,}298 - R/(n+0{\cdot}55)^2 \quad (n = 3,\ 4,\ 5,\ \text{etc}). \tag{2}$$

It is not possible to be quite as certain of all the terms of this series as it is of those of series (1), because there is considerable overlapping at its limit. The associated ionization potential is 10·027 V The difference between the two limits (i.e. 436 cm.$^{-1}$) is about the value which might be expected for the doublet separation of the $^2\Pi_g$ state of CS_2^+. The series (1) and (2) do not account for all the bands in the region 1550–1230 A. Other series must be present though they are difficult to establish It should be stressed that all the bands in this region appear to correspond to vibrationless electronic transitions, in agreement with the non-binding character of the electron removed. The predicted ionization potential for a completely non-bonding $3p_8$ electron is 10·8 V (Mulliken 1935*d*). The discrepancies with the observed values are possibly to be attributed in part to $\overset{-}{S}\overset{+}{C}\overset{-}{S}$ polarity, though this effect cannot be great as sulphur is actually below carbon on the electronegativity scale of the elements (Pauling 1932, Mulliken 1934). Most of the discrepancy must therefore arise from the fact that $(\pi-\pi, \pi_g)$ is slightly S⟷S anti-bonding. The slight increases of the ν_1 and ν_3 vibrations associated with the 1600 A band relative to their values in the ground state supports this.

TABLE V. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES FOR THE SERIES (1) AND (2) IN CARBON DISULPHIDE, TOGETHER WITH THE WAVE NUMBER SEPARATION OF THE CORRESPONDING MEMBERS

	Series (1)		Series (2)		
n	ν obs.	ν calc.	ν obs.	ν calc.	(1)–(2)
3	73,050	73,026	72,570	72,590	480
4	76,400	76,433	75,974	75,997	426
5	78,170	78,171	77,732	77,335	438
6	79,180	79,176	Obscured	78,740	—
7	79,810	79,809	79,370	79,393	440
8	80,236	80,230	(79,810)	79,797	426
9	80,515	80,531	80,087	80,095	428
10	80,743	80,748	80,307	80,312	436
11	80,911	80,911	80,475	80,475	436
12	81,036	81,037	80,590	80,601	446
13	81,136	81,137	(80,710)	80,700	426
14	81,218	81,215	80,778	80,780	438
∞	—	81,734	—	81,298	436

( ) means used in series (1), i.e. probably two overlapping bands. The accuracy of the measurements varies according to the nature of the band measured. For some of the sharper bands it is as great as 5 cm.$^{-1}$

Below 1200 A various sets of diffuse bands appear, each consisting of a fairly wide vibrational pattern. They are almost certainly due to the excitation of a π_u electron from the double bond. Judging from the wave-length at which they first appear (1200 A), they probably go to an ionization potential in the neighbourhood of 13·5 V (i.e. ~3·5 V more than for the non-bonding electron). This is compatible with the values assigned to the analogous electrons in carbon dioxide (Mulliken 1935*c*), and suggests that the CS_2^+ emission bands are to be expected in the region 3500 A. The diffuseness of the bands below 1200 A is probably due to pre-ionization. The excited π_u electron communicates its energy (~10·5 V) to the π_g

electron (e.g. by collision) and so ejects it. Absorption bands situated just above the minimum ionization potential of a molecule frequently show this phenomenon. Oxygen is a striking example, here direct photo-ionization of a $v\pi$ electron is very improbable (Price and Collins 1935), but the Hopfield bands ($w\pi$ electron excited), which lie just above the ionization potential of this electron (i.e. ~1000 A) show considerable diffuseness, which is almost certainly due to pre-ionization.

Before concluding it is perhaps worth while making some remarks in connexion with the electronic nature of the 3000 and 2000 A absorption regions of sulphur dioxide and carbon disulphide. Roughly speaking, it may be said that near ultra-violet absorption spectra correspond to transitions to anti-bonding orbitals associated with the electronic configuration of the ground state, or with configurations which differ but little from it. Far ultra-violet absorption spectra, on the other hand, are of a Rydberg type, and correspond to electronic jumps involving change in the principal quantum number. As the number of excited states associated with a given molecular configuration increases with the complexity of the bonds, it is natural that near ultra-violet absorption spectra should be most prolific for those molecules in which there is the greatest degree of unsaturation. Electrons in single bonds (e.g. C—C, C—H) do not give rise to any absorption in the near ultra-violet. They seem only to absorb in their first strong resonance (Rydberg) bands which lie in the vacuum region. Double bonds, on the other hand, possess two types of absorption. In addition to the absorption of the strong resonance bands, there is a much weaker absorption between 2000 and 2500 A (e.g. work by Carr and collaborators (1936) on ethylene and its derivatives) This has been attributed to the excited state $[x+x]^{-1}[x-x, b_{2g}]^1B_{1u}$. (Mulliken 1935*a*), i.e the transition of the outer bonding electron to the lowest anti-bonding orbital of the double bond. A similar explanation has been put forward by Sklar (1937) to account for the various near ultra-violet absorption regions of benzene The far ultra-violet absorption of benzene observed by Price and Wood (1935) corresponds to the photo-ionization of the π_h electrons of this molecule, the spectra being in excellent agreement with what is expected of these electrons.† The near ultra-violet absorption spectrum of benzene is shifted to the red relative to that of ethylene, because it corresponds to the excitation of a π_h electron from the more weakly bonding group. If an electron starting from a non-bonding orbital jumps into an anti-bonding orbital, then one would expect the absorption to appear still further to the red, since the electron jumps from a higher initial state. This is apparently what happens in the first ketonic absorption which occurs around 3000 A (for formaldehyde the upper state of this band is most probably $(2py)_{\mathrm{O}}^{-1}\,[z_{\mathrm{CH}_2}-z_{\mathrm{O}}]$ (Mulliken 1935*b*)). Now there is considerable apparent similarity between the absorption spectra of ketones and that of carbon disulphide and sulphur dioxide. This is no doubt to be related to the fact that in these molecules we have lone pairs adjoining double bonds. Thus it might be expected that the near ultra-violet bands of the latter molecules would correspond to the excitation of a non-bonding electron from the S or O atom respectively, to some anti-bonding orbital of the double bond. In addition to the similarity of the 3000 A systems, it seems that the 2000 A band in ketones bears considerable relation to the bands in carbon disulphide and sulphur dioxide, occurring in this particular region. There is much evidence to show that the bands in acetone, acetaldehyde and formaldehyde, which lie in the range 2000–1700 A, are different in nature from the Rydberg bands at shorter wave-lengths. They do not fit into the Rydberg formulae which express the latter bands and, further, in the case of acetone and acetaldehyde their appearance is quite different It is possible that they may be ascribed to a transition $2p \rightarrow (x_{\mathrm{C}}-x_{\mathrm{O}})$ Thus, if the corresponding bands in carbon disulphide and sulphur dioxide are similar in origin, it seems likely that this fact will help in the final identification of their excited states It must be stressed, however, that although relations can be traced between $>\mathrm{C}=\mathrm{O}$ and $\mathrm{O}=\mathrm{C}=\mathrm{O}$, or $>\mathrm{C}=\mathrm{S}$ and $\mathrm{S}=\mathrm{C}=\mathrm{S}$, and in both cases it is justifiable to speak of $\mathrm{C}=\mathrm{O}$ and $\mathrm{C}=\mathrm{S}$ double bonds, the electronic structure of these bonds is very different (Mulliken 1935*c*). It is the difference in the electronic nature of the bonding orbitals (i e. their closed shell configuration) which accounts for the so-called resonance energy of the CO bond in CO_2 relative to that in ketones (Pauling and Sherman 1933). The non-bonding electrons are apparently not much affected by the resonance. Also it should be pointed out that similarities between near ultra-violet spectra can only be drawn when the ionization potentials of the molecules compared do not differ greatly. For instance, the bands in CO_2 which are probably analogous to the 3000 A bands of CS_2, occur below 1700 A (Leifson 1926; Mulliken 1935*c*). This is due to the fact that the ionization potential of CO_2 is nearly 4 V greater than that of CS_2.

---

† The electrons from the "four" group should be very little bonding, and correspond to the lowest ionization potential, those of the "two" group should be more bonding, and go to the higher ionization potential.

In conclusion we wish to thank the Goldsmiths' Company (D. M. S.), and the Royal Society Grants Committee (W. C. P.), for financial aid in connexion with this work.

SUMMARY

The absorption spectra of sulphur dioxide and carbon disulphide have been investigated by means of a vacuum spectrograph down to 1000 A. For both molecules the systems of bands can be divided into two classes. (1) those which exhibit wide vibrational structure, (2) those which exhibit little or no vibrational structure. The former class probably correspond to transitions to anti-bonding molecular orbitals, while the latter are due to the transitions of comparatively non-bonding electrons to excited orbitals, which are mainly atomic in character. In the case of sulphur dioxide, the extrapolation of the bands of class (2) to their limit gives a value of $12{\cdot}05 \pm 0{\cdot}05$ V for the ionization potential of the molecule. A similar procedure for carbon disulphide yields the much more accurate values $10{\cdot}083$ and $10{\cdot}027$ V for ionization to the two components of the doublet state $^2\Pi_g$ of CS_2^+ ($\Delta\nu = 436$ cm.$^{-1}$). The experimental evidence indicates that while carbon disulphide is slightly bent in the earlier stages of the excitation, it finally returns to a linear configuration in CS_2^+. A vibrational analysis of the bands of class (1) is also given, and some general features of the electronic spectra of polyatomic molecules are discussed.

REFERENCES

Asundi and Samuel 1935 *Proc Ind. Acad. Sci.* **2**, 30–45.
Brockway 1936 *Rev. Mod. Phys.* **8**, 231–66.
Carr and Walker 1936 *J. Chem. Phys.* **4**, 751–55.
Carr and Walter 1936 *J. Chem. Phys.* **4**, 756–60.
Chow 1933*a* *Phys. Rev* **44**, 638–43
— 1933*b* *Chin. J. Phys.* **1**, 1–37.
Clements 1935 *Phys. Rev.* **47**, 224–32.
Collins and Price 1934 *Rev. Sci. Instrum.* **5**, 423–5.
Crawford and Huff 1936 *Phys. Rev.* **49**, 413 (A).
Franck, Sponer and Teller 1932 *Z phys. Chem* B, **18**, 88–102
Henri 1931 *Leipzige Vorträge*, pp. 131–54.
Herzberg and Teller 1933 *Z. phys. Chem.* B, **21**, 410–46.
Hutchisson 1930 *Phys. Rev.* **36**, 410–20.
— 1931 *Phys. Rev.* **37**, 45–52.
Jenkins 1929 *Astrophys. J.* **70**, 191–3.
Jonescu 1933 *C.R. Acad. Sci., Paris*, **196**, 1476–8.
Leifson 1926 *Astrophys. J.* **63**, 73–89.
Lotmar 1933 *Z. Phys.* **83**, 765–85.
Mulliken 1934 *J. Chem. Phys.* **2**, 782–93.
— 1935*a* *J. Chem. Phys.* **3**, 518–28.
— 1935*b* *J. Chem. Phys.* **3**, 564–73.
— 1935*c* *J. Chem. Phys.* **3**, 720–39
— 1935*d* *J. Chem. Phys.* **3**, 514–17.
— 1937 *J. Phys. Chem.* **41**, 159–73.

Pauling 1932 *J. Amer. Chem. Soc.* **54**, 3570–82
Pauling and Sherman 1933 *J. Chem. Phys.* **1**, 606–17.
Price 1936*a* *J Chem Phys.* **4**, 147–53.
— 1936*b* *J. Chem. Phys* **4**, 539–49.
Price and Collins 1935 *Phys. Rev.* **48**, 714–19.
Price and Wood 1935 *J. Chem. Phys* **3**, 439–44.
Rathenau 1933 *Z. Phys.* **87**, 32–56.
Sklar 1937 *J. Chem Phys.* **5**, 669–81.
Smyth and Blewett 1934 *Phys Rev.* **46**, 276–7.
Smyth and Mueller 1933 *Phys. Rev.* **43**, 121–2.
Watson and Parker 1931*a* *Phys. Rev.* **37**, 1013 (A).
— — 1931*b* *Phys. Rev.* **37**, 1484–92.
Wilson 1929 *Astrophys. J* **69**, 34–42

A suggestion for unifying quantum theory and relativity

By M. Born

(*Communicated by E T. Whittaker, F.R.S.—Received* 5 *January* 1938)

INTRODUCTION

There seems to be a general conviction that the difficulties of our present theory of ultimate particles and nuclear phenomena (the infinite values of the self energy, the zero energy and other quantities) are connected with the problem of merging quantum theory and relativity into a consistent unit. Eddington's book, "Relativity of the Proton and the Electron", is an expression of this tendency, but his attempt to link the properties of the smallest particles to those of the whole universe contradicts strongly my physical intuition. Therefore I have considered the question whether there may exist other possibilities of unifying quantum theory and the principle of general invariance, which seems to me the essential thing, as gravitation by its order of magnitude is a molar effect and applies only to masses in bulk, not to the ultimate particles. I present here an idea which seems to be attractive by its simplicity and may lead to a satisfactory theory.

1. RECIPROCITY

The motion of a free particle in quantum theory is represented by a plane wave

$$A\exp\left[\frac{i}{\hbar}p_k x^k\right],$$

where x^1, x^2, x^3, x^4 are the co-ordinates of space-time x, y, z, ct, and p_1, p_2, p_3, p_4 the components of momentum-energy p_x, p_y, p_z, E. The expression is completely symmetric in the two 4-vectors x and p. The transformation theory of quantum mechanics extends this "reciprocity" systematically. In a representation of the operators x^k, p_k in the Hilbert space for which the x^k are diagonal (δ-functions), the p_k are given by $\frac{\hbar}{i}\frac{\partial}{\partial x_k}$; and vice versa, if the p_k are diagonal the x^k are given by $-\frac{\hbar}{i}\frac{\partial}{\partial p_k}$. Any wave equation in the x-space can be transformed into another equation in the p-space, by help of the transformation

$$\phi(p) = \int \psi(x) \exp\left[\frac{i}{\hbar} p_k x^k\right] dx.$$

This reciprocity[*] can be extended also to the case of particles subject to external forces where the waves are not plane.

But there is a break in the reciprocal treatment when the principles of general relativity are applied. This theory has its origin in astronomical questions connected with the law of gravitation, and is founded on the conception of classical mechanics where the motion of a mass particle is represented not by a wave function, but by an orbit. The fundamental notion is the four-dimensional line element

$$ds^2 = g_{kl}\,dx^k dx^l, \tag{1}$$

the coefficients g_{kl} of which form the metric tensor.

It occurred to me that the principle of reciprocity would lead to the consideration of a line element in the p-space

$$d\sigma^2 = \gamma^{kl}\,dp_k dp_l \tag{2}$$

defining a metric in this space, but one which is not directly connected with the metric tensor g_{kl} in the x-space. If classical mechanics were valid throughout, this assumption would of course be impossible, for then p^k would be equal to $\mu\dot{x}^k$, where μ is the rest mass and the dot means differentiation with respect to proper time; therefore the transformation laws of the vector p would be completely determined by that of the vector x, and it would not be admissible to assume an independent absolute quadric for the determination of the metric in the p-space. But the real laws of nature are those of quantum theory. The classical conceptions refer only to a limiting case, namely, that which is apt to describe the motion of molar bodies in space-time. It is characterized by the condition that energy and momentum of the quanta involved ($h\nu$ and h/λ) are extremely small (as compared with $h\nu_0$ and h/λ_0, where $\lambda_0 = c/\nu_0$ is the Compton wave length), whereas space and time are unlimited. There is another possibility of going over to a limit, namely, the case where we have to do with very small regions of space and time (as compared with λ_0 and $1/\nu_0$), but with unrestricted amounts of energy and momentum. This is the domain of ultimate particles and nuclear phenomena. It seems to me unjustified to assume that these two reciprocal limiting cases should be subject to the same metric, based on the line element in the x-space. I suggest that the conception of a metric is inapplicable for those phenomena in which x-space and p-space are involved simultaneously with about equal weight, it is only valid for the two limiting cases, for molar processes in the x-space, and for nuclear processes in the p-space. I have the impression that this assumption does not contradict any known fact. We have learned that the simultaneous measurement of a co-ordinate x^k and a momentum p_k are restricted by the uncertainty laws (which, by the way, conform to the principle of reciprocity, as they contain the x^k and p_k symmetrically). They should provide for the freedom necessary to have different and widely independent metrics for the two limiting cases, which we shall call, for sake of brevity, the molar and the nuclear world.

---

[*] The word "reciprocity" is chosen because it is already generally used in the lattice theory of crystals where the motion of the particle is described in the p-space with help of the "reciprocal lattice".

2. THE DIFFERENTIAL EQUATIONS FOR THE METRIC TENSORS

In Einstein's theory of gravitation the metric tensor g_{kl} has to satisfy differential equations which connect the curvature tensor R_{kl} of space-time with the tensor energy-density T_{kl} of matter (including electromagnetic field). The most general form of these equations is

$$R_{kl}-(\tfrac{1}{2}R+\lambda)\,g_{kl} = -\kappa T_{kl}, \tag{3}$$

where κ is Einstein's gravitational constant and λ the cosmological constant. It is well known that these equations have a static solution corresponding to a closed (hyperspheric) world filled with matter of uniform density. Therefore there exists an upper limit for the distance between two points, given by the radius a of the universe.

Let us transfer this consideration to the p-space. For this purpose we have to define its curvature tensor P^{kl} in exactly the same way as the R_{kl} in the x-space. Further, we have to introduce quantities T^{kl} depending on the presence of matter. The meaning of these becomes clear if we remember that in the x-space the integrals $\int T_{k4}\,dx\,dy\,dz$ are momentum and energy of the system considered; analogously the integrals $\int \mathrm{T}^{k4}\,dp_x\,dp_y\,dp_z$ must be interpreted as space co-ordinates and time value of the system. We have, therefore, in accordance with our general considerations, to attribute to the whole system one single point in space (which may move in time); spatial specifications of the parts of the system are meaningless, whereas we have full freedom to study the energetic processes of the parts.

This seems to be a proper way of dealing with internuclear processes. As far as I can see the existing theories of the nuclei are of this type. For the description of the fundamental properties of a nucleus it seems to be unnecessary to specify carefully the law of interaction between its constituent particles, any function of the distance will do, if only the total range of action and the dissociation energy are properly chosen. The fully developed theory should, of course, modify the extreme p-standpoint and allow some statements on spatial properties of nuclei in accordance with the uncertainty rules

The differential equation for the metric tensor γ^{kl} in the p-space will have the same form as that in the x-space, namely

$$\mathrm{P}^{kl} - (\tfrac{1}{2}\mathrm{P} + \lambda')\gamma^{kl} = -\kappa'\mathrm{T}^{kl}, \tag{4}$$

where λ' and κ' are constants. Whether these nuclear constants are connected with the corresponding molar constants λ, κ cannot be decided yet.

3. HYPERSPHERICAL MOMENTUM SPACE

The equations (4) will have a solution corresponding to a closed (hyperspheric) momentum space (p_x, p_y, p_z), independent of E. Therefore we are led to the conclusion that for systems of some kind there is an *upper limit for momenta*,* determined by the radius b of the hypersphere. The systems to which this idea is applicable must be *energetically closed*; it certainly does not apply to every system, as we know the existence of particles with any amount of momentum and energy (cosmic rays).

This result is of great importance, as it removes immediately the infinities which are the dark points of the present theories. The hypersphere can be written by help of a parameter u, having the character of a momentum,

---

* This assumption has already been made, but without any relativistic foundation, by M Born and G. Rumer (1931). See also G. Wataghin (1934) and A. March (1937). Quite a different way of avoiding the infinite self-energy has been suggested by G. Wentzel (1933, 1934).

$$p_x^2+p_y^2+p_z^2+u^2 = b^2; \tag{5}$$

from this we get

$$u = (b^2-p^2)^{\frac{1}{2}}, \quad u\,du = -(p_x dp_x + p_y dp_y + p_z dp_z).$$

The line element of the p-space we get by eliminating u and du from

$$d\sigma^2 = dE^2 - (dp_x^2+dp_y^2+dp_z^2+du^2)$$

in the form

$$d\sigma^2 = dE^2 - \left\{dp_x^2\left(1+\frac{p_x^2}{b^2-p^2}\right)+dp_y^2\left(1+\frac{p_y^2}{b^2-p^2}\right)+dp_z^2\left(1+\frac{p_z^2}{b^2-p^2}\right)\right.$$
$$\left.+2dp_y dp_z\frac{p_y p_z}{b^2-p^2}+2dp_z dp_x\frac{p_z p_x}{b^2-p^2}+2dp_x dp_y\frac{p_x p_y}{b^2-p^2}\right\}. \tag{6}$$

We omit the well-known proof that the γ^{kl} defined by (6) are solutions of the differential equations (4) if b is suitably chosen as a function of λ', κ'.

The three-dimensional volume element is given by

$$d\Omega = \sqrt{(-\gamma)}\,dp_x dp_y dp_z,$$

where

$$-\gamma = \begin{vmatrix} 1+\dfrac{p_x^2}{b^2-p^2} & \dfrac{p_x p_y}{b^2-p^2} & \dfrac{p_x p_z}{b^2-p^2} \\ \dfrac{p_y p_x}{b^2-p^2} & 1+\dfrac{p_y^2}{b^2-p^2} & \dfrac{p_y p_z}{b^2-p^2} \\ \dfrac{p_z p_x}{b^2-p^2} & \dfrac{p_z p_y}{b^2-p^2} & 1+\dfrac{p_z^2}{b^2-p^2} \end{vmatrix} = 1+\frac{p^2}{b^2-p^2} = \frac{b^2}{b^2-p^2}.$$

This shows that b is the upper limit of p. We get

$$d\Omega = \frac{dp_x dp_y dp_z}{\sqrt{(1-p^2/b^2)}}. \tag{7}$$

This simple result admits of some important applications. For if we have to do with a system of independent particles, the fundamental law of quantum statistics gives the *number of quantum states of weight g in a spatial volume V and a momentum element $d\Omega$*

$$dn = g\frac{V}{h^3}d\Omega = g\frac{V}{h^3}\frac{dp_x dp_y dp_z}{\sqrt{(1-p^2/b^2)}}, \tag{8}$$

The appearance of the square root indicates deviations from the classical laws; it removes, as stated above, the disturbing infinities. We shall show this for a few examples connected with quantum electrodynamics.

The *total number of quantum states* in V is finite, namely,

$$n = \int dn = g\frac{V}{h^3}\iiint \frac{dp_x\,dp_y\,dp_z}{\sqrt{(1-p^2/b^2)}} = g\frac{4\pi Vb^3}{h^3}\int_0^1 \frac{\xi^2\,d\xi}{\sqrt{(1-\xi^2)}} = g\frac{\pi^2 Vb^3}{h^3}. \qquad (9)$$

The important question arises whether the constant b is universal, or characteristic for each energetically closed system. I do not think this question can be answered in the present preliminary state of the theory. For the sake of argument I shall assume in the following examples of application that the value of b is always the same.

4. APPLICATION TO QUANTUM ELECTRODYNAMICS

We use the form of quantized electrodynamics given by Fermi (1932). He writes the Hamiltonian for a system of electrons in an electromagnetic field contained in the volume V, as the operator

$$H = H_e + H_r + H_i + \frac{1}{\pi V}\sum_s \frac{h^2}{p_s^2}\Big(\sum_k e_k \cos\Gamma_{sk}\Big)^2; \qquad (10)$$

here

$$H_e = -\sum_k \{c(\vec{\alpha}_k \vec{p}_k) + m_k c^2 \beta_k\} \qquad (11)$$

is the Dirac Hamiltonian for the electrons ($k = 1, 2, \dots$) with rest mass m_k,

$$H_r = \sum_s \{\tfrac{1}{2}(p_{s1}^2 + p_{s2}^2) + 2\pi^2\nu_s^2(q_{s1}^2 + q_{s2}^2)\} \qquad (12)$$

the energy of the oscillators representing the radiation field,

$$H_i = \sum_k e_k c\sqrt{\frac{8\pi}{V}}\sum_s \{\vec{\alpha}_k(\vec{A}_{s1} q_{s1} + \vec{A}_{s2} q_{s2})\}\sin\Gamma_{sk} \qquad (13)$$

the interaction energy between electrons and radiation, where $\vec{A}_{s1}$, $\vec{A}_{s2}$ are two unit vectors orthogonal to one another and to the direction of propagation of the wave s, which, at the place of the electron k, has the phase

$$\Gamma_{sk} = \frac{2\pi}{h}(p_{sx}x_k + p_{sy}y_k + p_{sz}z_k) + \delta_s. \qquad (14)$$

This theory represents the facts of radiation marvellously, but it involves some infinities. The simplest of these are:

(1) The zero energy of radiation contained in H_r; for the stationary states one has

$$H_r = \sum_s h\nu_s(n_s + \tfrac{1}{2}). \qquad (15)$$

(2) The Coulomb self-energy of the electrons contained in the term given explicitly in (9), namely,

$$H_c = \frac{1}{\pi V}\sum_s \frac{h^2}{p_s^2}\Big(\sum_k e_k \cos\Gamma_{sk}\Big)^2. \tag{16}$$

All these formulae may possibly need modifications as a consequence of the p-metric. But I do not expect these alterations will be essential, and I shall suppose here that the only effect of the p-metric is that on the counting of quantum states.

With help of (8), where $g = 2$ corresponding to the two directions of polarization, the zero energy of radiation becomes

$$E_0 = H_r^0 = \sum_s \tfrac{1}{2}h\nu_s = \frac{c}{2}\sum_s p_s = \frac{8\pi V_c}{2h^3}\int_0^b \frac{p^3\,dp}{\sqrt{(1-p^2/b^2)}}$$
$$= \frac{4\pi c V b^4}{h^3}\int_0^1 \frac{\xi^3\,d\xi}{\sqrt{(1-\xi^2)}} = \frac{8\pi}{3}\frac{c V b^4}{h^3},$$

or with help of (9)

$$E_0 = \frac{4}{3\pi}cbn, \tag{17}$$

which has, in fact, the dimension of energy. The Coulomb interaction (16) can be written

$$H_c = \frac{h^2}{\pi V}\sum_{k,l} e_k e_l R_{kl}, \tag{18}$$

with

$$R_{kl} = \sum_s \frac{\cos\Gamma_{sk}\cos\Gamma_{sl}}{p_s^2} = \frac{V}{h^3}\iiint \cos\Gamma_k \cos\Gamma_l \frac{dp_x\,dp_y\,dp_z}{p^2\sqrt{(1-p^2/b^2)}}, \tag{19}$$

where the weight g has to be taken equal to 1 (longitudinal waves).

We average over all phases δ_s and introduce the cosine γ of the angle between the vectors $\vec{p}$ and $\vec{r}_{kl} = \vec{r}_k - \vec{r}_l$. Then

$$R_{kl} = \frac{2\pi V}{2h^3}\int_0^b\int_{-1}^1 \cos\left(\frac{2\pi}{h}p\gamma r_{kl}\right)\frac{dp\,d\gamma}{\sqrt{(1-p^2/b^2)}}$$
$$= \frac{V}{h^2}\int_0^b \frac{\sin\left(\frac{2\pi}{h}pr_{kl}\right)}{pr_{kl}}\frac{dp}{\sqrt{(1-p^2/b^2)}}.$$

If we introduce the function

$$f(x) = \frac{2}{\pi}\int_0^1 \frac{\sin(\xi x)}{\xi}\frac{d\xi}{\sqrt{(1-\xi^2)}} = \int_0^x J_0(y)\,dy, \tag{20}$$

where $J_0(y)$ is the Bessel function, we get

$$R_{kl} = \frac{\pi V}{2h^2} \frac{1}{r_{kl}} f\left(2\frac{r_{kl}}{r_0}\right), \tag{21}$$

with

$$r_0 = \frac{h}{\pi b}. \tag{22}$$

Substituting (21) in (18) we get the modified Coulomb law

$$H_c = \tfrac{1}{2} \sum_{k,l} \frac{e_k e_l}{r_{kl}} f\left(2\frac{r_{kl}}{r_0}\right). \tag{23}$$

From the definition (20) one finds easily

$$\underset{x \to \infty}{\mathrm{L}} f(x) = 1, \quad \underset{x \to 0}{\mathrm{L}} \frac{1}{x} f(x) = 1. \tag{24}$$

Therefore we have the classical Coulomb law for $r_{kl} \gg r_0$, and r_0 determines evidently the "dimensions" of the electron. One finds its precise meaning by calculating the self-energy terms in (23), taking $k = 1$, namely,

$$\frac{e^2}{r_0} = mc^2, \tag{25}$$

where we have introduced the mass m of the electron. Therefore r_0 is the classical radius of the electron, $r_0 = e^2/mc^2 = 2{\cdot}80 \times 10^{-13}$ cm., and we get from (22)

$$b = \frac{h}{\pi r_0} = 7{\cdot}43 \times 10^{-15}\,\mathrm{g.\,cm.\,sec.}^{-1}. \tag{26}$$

As the terms H_c account for the inertia of the electrons one would be inclined to omit the mass terms in the Dirac Hamiltonian H_e (11); but these appear there multiplied by the spin operator β. This shows that a complete explanation of mass as an electromagnetic phenomenon requires a deeper understanding of the relation of the spin and the electromagnetic field. We shall not go into this question.

Introducing r_0 from (22) instead of b in (9) and (17), one gets

$$n = n_0 V, \quad n_0 = \frac{2}{\pi r_0^3} = 2{\cdot}90 \times 10^{37}\,\mathrm{cm.}^{-3}; \tag{27}$$

$$E_0 = \epsilon_0 n = \epsilon_0 n_0 V, \quad \epsilon_0 = \frac{4}{3\pi^2}\frac{ch}{r_0} = \frac{8}{3\pi}\frac{hc}{2\pi e^2}\frac{e^2}{r_0} = \frac{8}{3\pi} 137\, mc^2$$

$$= 117\, mc^2 = 9{\cdot}49 \times 10^{-5}\,\mathrm{erg} = 5{\cdot}97 \times 10^7\,\text{e-volts};$$

$$\epsilon_0 n_0 = \frac{16}{3\pi^2}\frac{h^2}{2\pi e^2}\left(\frac{e}{r_0^2}\right)^2 = \frac{16}{3\pi^2} 137 \left(\frac{e}{r_0^2}\right)^2 = 2{\cdot}75 \times 10^{33}\,\mathrm{erg\,cm.}^{-1}. \tag{28}$$

This shows that the zero energy per quantum oscillator, ϵ_0, is $8/3\pi \times 137$ times the rest energy of the electron, mc^2, and that the density of the zero energy of radiation, $E_0/V = \epsilon_0 n_0$, is $128/3\pi \times 137$ times the electrostatic energy density $\mathscr{E}^2/8\pi$, where $\mathscr{E} = e/r_0^2$ is the electric field at the "surface" of the electron.

The numerical values should be considered as preliminary, since the electromagnetic or transverse self-energy which arises from the term H_t, (13), has to be added. Dirac's single electron theory gives, according to Heitler (1936), for this transverse self-energy an expression which would lead to a value about 2×137 times as large as the electrostatic one. But it has been shown by Weisskopf (1934) and Kemmer (1935) that the hole theory of the electron leads to another expression which gives a value of the same order as the electrostatic one, differing only by a numerical factor $\frac{1}{2}$. In connexion with this question it should be considered whether the value of b for the longitudinal waves (electrostatic terms) and the transversal waves (electromagnetic terms) is necessarily identical

5. HEAT RADIATION

We can now apply formula (8) to the excited states of the radiation field. As the partition function of an oscillator with the energy $h\nu n$ is

$$Q_\nu = \sum_{n=0}^{\infty} e^{-h\nu n/kT} = \frac{1}{1-e^{-h\nu/kT}}, \tag{29}$$

we find for the free energy of the radiation field

$$F = -kT\sum_\nu \ln Q_\nu = kT\sum_\nu \ln(1-e^{-h\nu/kT}), \tag{30}$$

and if we assume that $p = h\nu/c$, we have with the help of (8)

$$F = \frac{8\pi kT}{c^3} V \int_0^{1/\tau} \ln(1-e^{-h\nu/kT}) \frac{\nu^2\,d\nu}{\sqrt{(1-(\nu\tau)^2)}}, \tag{31}$$

with
$$\tau = \frac{h}{bc} = \frac{\pi r_0}{c} = 2{\cdot}94 \times 10^{-23}\,\text{sec.}; \tag{32}$$

here we have assumed that b is the same as determined above; then τ is the time which light needs to travel the distance πr_0. The entropy is

$$S = -\frac{\partial F}{\partial T} = -\frac{F}{T} + \frac{8\pi h}{c^3} V \frac{1}{T} \int_0^{1/\tau} \frac{\nu^3\,d\nu}{(e^{h\nu/kT}-1)\sqrt{(1-(\nu\tau)^2)}} \tag{33}$$

and the energy
$$U = F+TS = V\int_0^{1/\tau} u(\nu, T)\,d\nu \tag{34}$$

with
$$u(\nu, T) = \frac{8\pi h}{c^3}\,\frac{\nu^3}{(e^{h\nu/kT}-1)\sqrt{(1-(\nu\tau)^2)}}. \tag{35}$$

This is the modified Planck formula for the density of radiation. The radiation pressure is given by

$$\begin{aligned} P = -\frac{\partial F}{\partial V} = -\frac{F}{V} &= -\frac{8\pi kT}{c^3}\int_0^{1/\tau}\ln(1-e^{-h\nu/kT})\,\frac{\nu^2 d\nu}{\sqrt{(1-(\nu\tau)^2)}} \\ &= -n_0 kT\frac{4}{\pi}\int_0^1 \ln(1-e^{-\Theta_0\xi/T})\,\frac{\xi^2 d\xi}{\sqrt{(1-\xi^2)}}, \end{aligned} \tag{36}$$

where

$$\left.\begin{aligned} k\Theta_0 = \frac{h}{\tau} = bc = \frac{hc}{\pi r_0} = \frac{3\pi}{4}\epsilon_0 = 2{\cdot}23\times 10^{-4}\,\text{erg} = 1{\cdot}41\times 10^{8}\,\text{e-volts}, \\ \Theta_0 = 1{\cdot}63\times 10^{12}\,\text{degrees}. \end{aligned}\right\} \tag{37}$$

These numerical values should be considered with reserve, as mentioned above.

The total energy density can be written

$$\begin{aligned} u(T) = \int_0^{1/\tau} u(\nu, T)\,d\nu &= n_0 k\Theta_0\frac{4}{\pi}\int_0^1 \frac{\xi^3 d\xi}{(e^{\Theta_0\xi/T}-1)\sqrt{(1-\xi^2)}} \\ &= n_0 k\Theta_0\frac{4}{\pi}\int_0^{\pi/2}\frac{\sin^3\phi\, d\phi}{e^{\Theta_0 \sin\phi/T}-1}. \end{aligned} \tag{38}$$

The quantity $u-3P$ which vanishes in the classical theory differs here from zero, namely

$$u(T)-3P(T) = n_0 kT\frac{4}{\pi}\int_0^1 \frac{\xi^2 d\xi}{\sqrt{(1-\xi^2)}}\left\{\frac{\xi\Theta_0/T}{e^{\xi\Theta_0/T}-1}+3\ln(1-e^{-\Theta_0\xi/T}\right\}. \tag{39}$$

This vanishes for $T\to 0$, but has for $T \gg \Theta_0$ the value $n_0 kT$, corresponding to a kind of saturation as if each degree of freedom of the vacuum had acquired the equipartition value kT of energy. As a matter of fact, the formula (38) for the energy is formally similar to that of a (one-dimensional) crystal lattice, as studied a long time ago by v. Kármán and myself (1912, 1913). One has for $T \ll \Theta_0$ the Stephan-Boltzmann law

$$u(T) = aT^4, \quad a = \frac{8\pi^5}{15}\frac{n_0 k}{\Theta_0^3} = \frac{8\pi^5 k^4}{15 c^3 h^3}, \tag{40}$$

whereas for $T \gg \Theta_0$
$$u(T) = n_0 kT; \tag{41}$$

in this region the vacuum follows the law of Dulong-Petit.

These results show that the radiation pressure cannot be considered as the transfer of momentum $p = h\nu/c$. This holds only for temperatures low compared with Θ_0; for higher temperatures the pressure has more the character of the internal pressure of a vibrating crystal lattice. It follows that Maxwell's equations cannot hold for high-frequency waves, but have to be modified in such a way that the relation of the pressure of light to the density of energy is consistent with (39). But as Fermi's formulae from which we started are nothing but the quantized Maxwell's equations there may possibly be deeper alterations necessary affecting all the formulae of this section.

6. KINETIC THEORY OF GASES

There are also deviations from the accepted laws of the kinetic theory of gases. The partition function per molecule becomes*

$$Q = \frac{V}{h^3}\iiint e^{-p^2/2MkT}\frac{dp_x\,dp_y\,dp_z}{\sqrt{(1-p^2/b^2)}}$$

or
$$Q = Vn_0\frac{2}{\pi}\int_0^1 e^{-\Theta\xi^2/T}\frac{\xi^2\,d\xi}{\sqrt{(1-\xi^2)}} = Vn_0\tfrac{1}{2}e^{-\Theta/2T}\left\{J_0\left(\frac{i\Theta}{2T}\right)+iJ_1\left(\frac{i\Theta}{2T}\right)\right\}, \tag{42}$$

where Θ can be expressed by the characteristic temperature of the vacuum:

$$\Theta = \frac{b^2}{2Mk} = \frac{m}{M}\frac{hc}{2\pi e^2}\Theta_0 = \frac{137}{1845\mu}\Theta_0 = \frac{1{\cdot}21\times 10^{11}}{\mu}\text{ degree.} \tag{43}$$

μ is the molecular weight relative to the H-atom. One has for $T \ll \Theta$ the usual formula

$$Q = V\frac{1}{2\sqrt{\pi}}n_0\left(\frac{T}{\Theta}\right)^{\frac{3}{2}} = V\left(\frac{2\pi MkT}{h^2}\right)^{\frac{3}{2}}, \tag{44}$$

but for $T \gg \Theta$

$$Q = \tfrac{1}{2}Vn_0\left\{1-\frac{3}{4}\frac{\Theta}{T}+\frac{5}{16}\left(\frac{\Theta}{T}\right)^2-\frac{35}{384}\left(\frac{\Theta}{T}\right)^3+\frac{21}{1024}\left(\frac{\Theta}{T}\right)^4-\dots\right\}. \tag{45}$$

This high temperature degeneration has no influence on the equation of state, but on the specific heat. It may play a role in the theory of the constitution of stars, and on the constitution of nuclei as well, as these, according to Bohr and Kalckar (1937), can be treated by thermodynamical methods.

---

* This problem can also be treated relativistically quite easily; however, the result depends essentially on what assumption one makes about the radius b.

The molar heat for high temperatures, $T \gg \Theta$, is

$$c_v = \frac{R}{16}\frac{\Theta^2}{T^2} + \dots, \tag{46}$$

and goes to zero for $T \to \infty$.

CONCLUSION

A consequence of the assumption of a finite size of a system in the p-space is the existence of a set of proper functions $\psi_n(p)$, where the index n refers to proper values of some functions of the space co-ordinates. This means that our theory leads to a kind of granular or lattice structure of space without introducing such a strange assumption a priori.

The suggestions made in this paper contain an ample programme for further investigation, the most important question seems to me the generalization of the idea of the metric tensor and of the equations determining it, for that intermediate region where classical methods neither in the x-space nor in the p-space are applicable.

SUMMARY

The fact that the fundamental laws of quantum mechanics are symmetrical in space-time x^k and momentum-energy p_k can be generalized to a "principle of reciprocity" according to which the x-space and the p-space are subject to geometrical laws of the same structure, namely a Riemannian metric. In analogy with Einstein's closed x-world one has to assume that energetically closed systems (as elementary particles, nuclei) must be described by help of a hyperspherical p-space. A consequence of this assumption is a modification of the formula for the number of quantum states in an element of the p-space. The application of this formula to quantum electrodynamics leads to a finite zero energy of the vacuum, a finite self-energy of the electron, etc. Deviations from Planck's law and the Stephan-Boltzmann law of radiation, and the calorie properties of gases are predicted for very high temperatures.*

REFERENCES

Bohr, N. and Kalckar, F. 1937 *Kgl. Danske Vidensk. Selskab.*, Math.-fys. Med., **14**, 10. See further V. Weisskopf, *Phys. Rev.*, **52**, 295 (1937).

Born, M. and Kármán, Th. v. 1912 *Phys. Z.* **13**, 297

—— 1913 *Phys. Z.* **14**, 65.

---

* [*Note added in proof.*] The application of this theory to nuclei leads to results confirming the assumptions. Cf. *Nature*, **141**, 327 (1938).

Born, M. and Rumer, G. 1931 *Z. Phys.* **69**, 141.
Fermi, E. 1932 "Quantum theory of Radiation", *Rev. Mod. Phys.* **4**, 87; see especially formula (166), p. 130.
Heitler, W. 1936 "The Quantum theory of Radiation", Chap III, § 13, p. 183. (Oxford University Press.)
Kemmer, N. 1935 *Ann. Phys., Lpz.* (5), **22**, 674.
March, A. 1937 *Z. Phys.* **104**, 93, 161, **105**, 620, **106**, 49, 291, 532; **108**, 128.
Wataghin, G. 1934 *Z. Phys.* **88**, 92.
Weisskopf, V. 1934 *Z Phys.* **89**, 27; **90**, 817.
Wentzel, G. 1933 *Z Phys.* **86**, 479, 635.
— 1934 *Z. Phys.* **87**, 726.

Hyperfine structure, Zeeman effect and isotope shift in the resonance lines of potassium

By D. A. Jackson and H. Kuhn
Clarendon Laboratory, Oxford

(*Communicated by F. A. Lindemann, F.R.S.—Received* 24 *January* 1938)

[Plate 5]

Introduction

In an earlier work (Jackson and Kuhn 1935, 1936) the hyperfine structure in the resonance lines of the abundant isotope 39 of potassium was observed by the method of absorption in an atomic beam; but no intensity measurements were made. Qualitatively, the short wave-length component appeared to be the stronger, which led to the assumption of a negative magnetic moment. Magnetic deflexion experiments (Millman 1935; Fox and Rabi 1935), though in accurate agreement as regards the width of splitting, gave a positive magnetic moment. The absorption experiments were therefore repeated under conditions which excluded overlapping of neighbouring orders of the interferometer spectrum and thus permitted a quantitative determination of the intensities. This was achieved by using an etalon of 5 cm. length only (instead of 10 cm. in the old experiment) and by running the light source at low pressure of potassium. The measurements, the main results of which were published in a preliminary note (Jackson and Kuhn 1937*a*), gave an intensity ratio 1·45 of the hyperfine structure components, the long wave-length one being the stronger.

The nuclear spin of $\frac{3}{2}$ which follows from the magnetic deflexion experiments would require an intensity ratio of 1·66, which is definitely outside the limits of error of the intensity measurements. It was also observed that at high density of the atomic beam the component of shorter wave-length appeared wider than the other component. Both these facts were explained quantitatively by assuming that the lines of the 14 times rarer isotope K 41 overlapped the weaker component of K 39. This assumption is in agreement with the comparatively small splitting of the ground-level of K 41 found by the magnetic deflexion method (Manley 1936), and also explains why in the earlier photograms the component of shorter wave-length appeared stronger.*

In the following work it has been possible to prove the correctness of this assumption. By using three atomic beams of very high collimation, the lines of K 41 could be resolved from those of K 39

The hyperfine structure of the Zeeman effect of the resonance line $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$, $\lambda\,7664$ of the abundant isotope K 39 was observed by means of a single atomic beam. Each of the two π components was found to consist of four lines, proving conclusively that the value $I = \frac{3}{2}$ is correct.

EXPERIMENTAL

Light source. The potassium resonance lines $\lambda\,7664$ and $\lambda\,7699$ used as background for the absorption experiment were produced in an electrodeless discharge tube, filled with helium at a pressure of a few mm. A side tube containing metallic potassium was fitted to the capillary. It was heated in an electric furnace to a temperature of about 180° C. Under these conditions the resonance lines were strong and sufficiently free from self-reversal.

Atomic beam. The condition for resolving two lines of equal intensity at distance $\Delta\nu$ from each other is

$$h \leqq \Delta\nu,$$

h being their half-value width due to Doppler effect. The corresponding condition for resolving two lines of intensities i and I ($i < I$) is that the intensity of the strong line, at the distance $h/2$ from the weak one, must have fallen to at least the value $\frac{1}{2}i$. If $I/i = 2^x$, this condition is fulfilled if

$$h(\tfrac{1}{2}x + 1) \leqq \Delta\nu.$$

---

* In reality, it was equally strong, due to nearly complete absorption, but wider. With any shortage of resolving power, either of the etalon or the photometer, the wider of two equally strong lines appears slightly stronger.

For the present problem of resolving the lines of the rare isotope the factor in brackets is about 3, requiring $h \leq \frac{1}{3}\Delta\nu$. As the splittings to be resolved were expected to be about $\frac{1}{10}$ of the normal Doppler width, the condition required a collimation of the atomic beam of at least $\frac{1}{30}$.

In the actual experiment the collimation was made $\frac{1}{35}$. In order to obtain sufficient absorption for the rare isotope with this high degree of collimation, three atomic beams in series had to be used.

The atomic beams were formed in three vertical pyrex tubes of diameter 15 mm. with two linear constrictions of 5 mm width, at 17·5 cm. distance from each other. The absorption chamber to which these tubes were connected had plane windows sealed on and was evacuated through a wide tube leading to the pump. The potassium in the bottom ends of the atomic beam tubes was heated by means of three electric furnaces to temperatures between 150 and 220° C., corresponding to vapour pressures between 5×10^{-4} and 10^{-2} mm. The deposits on the top of the absorption chamber had extremely sharp edges, proving the absence of any appreciable amount of scattering.

For the investigation of the Zeeman effect of K 39 a single atomic beam tube was used, of collimation $\frac{1}{25}$. It was placed between the two cylindrical pole pieces of an electromagnet, of diameter 5 cm and separation 3·5 cm. A Nicol prism was introduced between the atomic beam tube and the spectrograph so that only the π components were transmitted.

Spectrograph and interferometer. A 1·5 m. spectrograph described in earlier work was used, with a very dense glass prism of 65°.

The interferometer was a Fabry-Pérot etalon with glass plates of 11 cm. diameter and separating pieces of silica. The plates were silvered by evaporation, the transmission of each being 4%. Separations of 5, 8 and 11 cm. were used.

The etalon was enclosed in an airtight box with plane glass windows. The box was fitted with an oil manometer and the air pressure inside could be varied by ± 10 cm. of oil. This arrangement has the advantage that slight variations in temperature or atmospheric pressure during an exposure do not alter the density of the air in the box. It is also possible, by adjusting the air pressure, to alter the position of the fringes within a range of one order. It was thus possible to bring the absorption lines near to the centre of the fringe system, where on account of the greater scale the resolution is highest.

The etalon was adjusted with the red ray of a cadmium lamp, the air pressure being set so that a small black spot appeared in the centre of the fringe system; the size of this spot is very sensitive to changes in thickness of the etalon. The whole silvered area of the plates (9 cm. in diameter) was used for adjustment, while for making the exposures the etalon was stopped down to 2 or 3 cm. diameter. Only in this manner was it possible to adjust the etalon with 11 cm. plate separation by means of the light of the cadmium red line, on account of the diffuseness of the fringes with this length of path.

RESULTS

Intensity ratio of the components of K 39

Fig. 1, Plate 5, shows an enlarged photograph and a photometer tracing of the line $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ taken with a 5 cm. etalon and a single atomic beam of collimation $\frac{1}{8}$. The light source shows a small amount of self-reversal. This condition is very suitable for intensity measurements, for the intensity curve of the background at the position of the absorption is very flat. The wide gaps between neighbouring orders show the absence of overlapping, which is an essential condition for determinations of intensities. The measurements gave the following values for the intensity ratio of the hyperfine structure components·

$4S_{\frac{1}{2}}$–$4\,^2P_{\frac{3}{2}}$ 1·44, 1·44, 1·44, 1·38, 1·43, 1·48, 1·45· mean 1·44;

$4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ 1·52, 1·42, 1·40, 1·45: mean 1·45.

In both lines the component of longer wave-length is the stronger. The method of measuring these intensities was that described in earlier work (Jackson and Kuhn 1937*b*). The correction for the overlapping of the Doppler wings of the two lines is negligible.

The value found agrees well with the intensity ratio which would be expected for a spin $\frac{3}{2}$ of K 39 if all the absorption of K 41 overlapped the weaker component of K 39, the value in this case being

$$1{\cdot}67/(1+2{\cdot}67/14) = 1{\cdot}40.$$

The observed value 1·45 is somewhat greater; this is to be expected as the lines of K 41 do not exactly overlap the weaker line of K 39 but lie on either side of it. As the magnetic deflexion experiments and the result of the investigation of the Zeeman effect (see p. 310) have established the value $I = \frac{3}{2}$ of K 39 beyond any doubt, this result is of interest as a test of the method of measuring intensities.

Structures of the lines of K 39 *and* K 41

Structure of the line $S_{\frac{1}{2}}$–$^2P_{\frac{3}{2}}$, λ7664. With an 8 cm. etalon, the triple atomic beam photographs showed, at low densities, the hyperfine structure doublet discussed above. At very much higher densities, a new line appeared on the long wave-length side of the weaker component of the doublet, and also a wing on the short wave-length side.

With an 11 cm. etalon this wing became a clearly resolved line, the $S_{\frac{1}{2}}-{}^2P_{\frac{3}{2}}$ line now showing a hyperfine structure pattern of four lines, as shown in the diagram fig. 4. Two of them, *a* and *b*, are only visible at densities of the atomic beam 10 or 20 times the density required to show the lines *A* and *B*, so that the assignment of *a* and *b* to the 14 times rarer isotope 41 is certain.

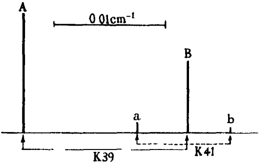


FIG. 4. Hyperfine structure of $4S_{\frac{1}{2}}-4\,{}^2P_{\frac{3}{2}}$

Fig. 2, Plate 5, shows an enlargement and a photometer curve of this structure.* A great number of photographs was taken; the four best plates were measured and gave the following results for the relative positions of the four components.

TABLE I. STRUCTURE OF $4S_{\frac{1}{2}}-4\,{}^2P_{\frac{3}{2}}$ (IN CM.$^{-1}$)

	A (K 39)	*a* (K 41)	*B* (K 39)	*b* (K 41)
	0·0000	0·0099	0·0150	0·0189
	0·0000	0·0104	0·0146	0·0184
	0·0000	0·0104	0·0148	0·0190
	0·0000	0·0101	0·0147	0·0182
Mean	0·0000	0·0102	0·0147	0·0186 cm.$^{-1}$

The line $S_{\frac{1}{2}}-{}^2P_{\frac{1}{2}}$, λ7699, was found to have a very similar structure (see diagram fig. 5). The component *b* is not completely resolved, *A* is much broader than in the other resonance line, and the beginning of a resolution into two lines can be seen on the original plates The results of measurements of four plates are given in Table II.

---

* In figs. 2 and 3 of Plate 5, as in most photographs taken with the 11 cm. etalon, the definition is better in one half of the fringe system—the right half in the figures—than in the other. This effect probably arises from small distortions of the etalon plates. The separation of 11 cm. of the etalon plates was obtained by putting together three separating pieces of length 5, 4 and 2 cm. On account of this a higher pressure than usual of the adjustment springs was needed, and this may have caused small distortions in the plates.

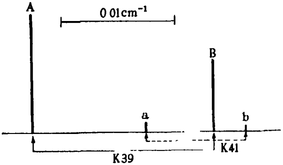


Fig. 5. Hyperfine structure of $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$.

TABLE II. STRUCTURE OF $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ (IN CM.$^{-1}$)

	A (K 39)	*a* (K 41)	*B* (K 39)	*b* (K 41)
	0·0000	0 0100	0 0156	0·0183
	0 0000	0 0103	0 0159	0·0190
	0·0000	0·0095	0·0160	0 0190
	0·0000	0 0104	0 0160	0·0191
Mean	0 0000	0 0100	0·0159	0·0188 cm.$^{-1}$

The smallest splitting resolved in these experiments is 0·0029 cm.$^{-1}$ ($B-b$ of $S_{\frac{1}{2}}$–$^2P_{\frac{1}{2}}$) which is $\frac{1}{11}$ of the normal Doppler half-value width at 200° C., the intensity ratio being 14/1. Applying the condition given above (p 304) the Doppler width which would just allow the resolution of these lines is found to be $\frac{1}{38}$ of the normal Doppler width at 200° C. Without the use of atomic beams, a temperature of less than 0·5° abs. would be required to resolve the lines.

The hyperfine structure of the Zeeman effect of the line
$4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ *of* K 39

The Zeeman effect of a line $S_{\frac{1}{2}}$–$^2P_{\frac{1}{2}}$ possesses two π components separated by $\frac{1}{3}$ of the total Lorentz splitting. A field strength was chosen at which the distance between the two π components was nearly equal to $\frac{1}{2}$ of an order of the etalon ($=0{\cdot}0227$ cm.$^{-1}$). The strength of this field was measured and found to be 730 gauss.

Each of the two π components was found to consist of four lines (fig. 3, Plate 5). On account of the rapid decrease in scale with the distance from the centre of the fringe system, the resolution is better in the long wave-length component.

In Table III are given the positions of the eight observed lines, zero being the position of the centre of gravity of the two components of the hyperfine structure without field, an intensity ratio of 5/3 being assumed. The accuracy of the measurements is ± 0·0005 cm.$^{-1}$.

TABLE III. ZEEMAN EFFECT OF $4S_{\frac{1}{2}}$-$4\,^2P_{\frac{1}{2}}$ (IN CM.$^{-1}$) π COMPONENTS

	−				+			
	0·0174	0·0137	0·0100	0·0064	0·0061	0·0107	0·0151	0·0181
	0·0170	0·0136	0·0100	0·0062	0·0052	0·0105	0·0146	0·0179
Mean	0·0172	0·0136	0·0100	0·0063	0·0056	0·0106	0·0148	0·0180

DISCUSSION OF THE RESULTS

K 39. *Hyperfine structure without magnetic field*

The resolution of the lines of the two isotopes made it possible to measure the structure of the lines of K 39 with higher accuracy than that obtained in the earlier work. The hyperfine structure splitting of the resonance lines is mainly due to the splitting of the ground level $4S_{\frac{1}{2}}$. The difference between the splittings of the two resonance lines is caused by the small, unresolved structures of the $4\,^2P$ terms Assuming that the measurements give the centres of gravity of the blends, from the separations 0·0147 and 0·0159 cm^{-1} of the components of the lines $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ and $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ it follows that

$$\Delta S_{\frac{1}{2}} = 0{\cdot}0153 \pm 0{\cdot}0002 \text{ cm.}^{-1}, \quad \Delta\,^2P_{\frac{1}{2}} = 0{\cdot}0018 \pm 0{\cdot}0006 \text{ cm}^{-1}.$$

The value of the splitting of the $4S_{\frac{1}{2}}$ term is practically the same as the value 0·0152 found in the first work (Jackson and Kuhn 1935, 1936) and agrees well with the magnetic deflexion measurements of Millman (1935), Fox and Rabi (1935), who found 0·0152 and 0·0154 cm.$^{-1}$. It leaves the value 0·39 of the magnetic moment unaltered.

The value $\Delta\,^2P_{\frac{1}{2}} = 0{\cdot}0018$ cm.$^{-1}$ is in agreement with estimates of the scarcely resolved splitting of the line A of $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$. It is also in agreement with recent experiments of Meissner and Luft (1937), who observed the hyperfine structure in the potassium resonance lines by means of an atomic beam of potassium excited by electron impacts. Nevertheless, too much importance must not be ascribed to the accuracy of Meissner's and Luft's measurements, as they did not resolve the lines of the weak isotope and made no allowance for possible displacements by them.

The nuclear magnetic moment, calculated from the $^2P_{\frac{1}{2}}$ splitting by means of Goudsmit's formula, is 0·37, in good agreement with the (more accurate) value derived from the $4S_{\frac{1}{2}}$ splitting.

K 39. *Zeeman effect*

Of the methods of determining nuclear spins, the most direct and reliable is the observation of the Zeeman effect. If the magnetic splitting is large compared with the hyperfine structure, each Zeeman component is split into $2I+1$ hyperfine structure lines (Goudsmit and Back 1927). For very great fields the lines are equidistant and of equal intensity; for fields that are smaller, but nevertheless high enough to give a Zeeman splitting considerably wider than the hyperfine structure, the number of components is still $2I+1$, but they are not quite uniformly spaced. For the line $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$ of K 39 the field strength of 730 gauss corresponds to this state, the total Lorentz-splitting being about 5 times greater than the hyperfine structure.

The position of the components can be calculated from the quantum-theoretical formulae (Goudsmit and Bacher 1930). Assuming the value $I=\frac{3}{2}$ of the nuclear spin and the splitting of 0·0153 cm.$^{-1}$ of the $4S_{\frac{1}{2}}$ term, the displacements (in cm.$^{-1}$) of the π components from the centre of gravity of the lines without field are given by

$$\delta\nu = \frac{0{\cdot}0153}{8} \mp \frac{0{\cdot}0153}{2}(1+mx+x^2)^{\frac{1}{2}} \pm \tfrac{1}{3}H \times 4{\cdot}7 \times 10^{-4},$$

where

$$x = \frac{e}{m} \times \frac{H}{4\pi c} \times \frac{g(J)}{0{\cdot}0153},$$

m is equal to m_J+m_I and has the values 2, 1, 0, -1 for the negative values of the second term, and 1, 0, -1, -2 for the positive values; and $g(J)$ is equal to 2.

The displacements calculated for a field of 730 gauss are given in the diagram of fig. 6, together with the observed values. It can be seen that they are in very good agreement, the observed closer spacing of the outer components, and also the values of the total widths of both groups are in very good agreement with theory. The effect of the unresolved structure of the term $4\,^2P_{\frac{1}{2}}$ on the Zeeman pattern is negligible.

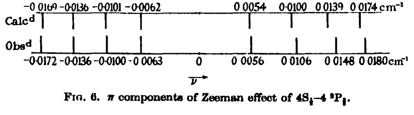


FIG. 6. π components of Zeeman effect of $4S_{\frac{1}{2}}$–$4\,^2P_{\frac{1}{2}}$.

K 41

The two lines *a* and *b* of the isotope 41 correspond to the lines *A* and *B* of 39. It follows that the magnetic moment of K 41 is positive, as the long wave-length component is the stronger one. The ratio of the splittings of K 39 to K 41 is

$$\frac{0{\cdot}0147}{0{\cdot}0084} = 1{\cdot}75 \text{ for the line } S_{\frac{1}{2}}-{}^2P_{\frac{3}{2}}, \text{ and}$$

$$\frac{0{\cdot}0159}{0{\cdot}0088} = 1{\cdot}80 \text{ for the line } S_{\frac{1}{2}}-P_{\frac{1}{2}}.$$

$$\text{Average } 1{\cdot}77 \pm 0{\cdot}05.$$

This agrees well with the value 1·80 found by Manley (1936) from magnetic deflexion experiments. Assuming the spin $I=\frac{3}{2}$ for K 41 found by Manley, the magnetic moment is found to be 0·22 nuclear magneton. The ratio of the magnetic moments of the two isotopes $\frac{\mu_{39}}{\mu_{41}} = 1{\cdot}77 \pm 0{\cdot}05$, being independent of the accuracy of the theoretical formulae, is much more accurate than their absolute values.

From the positions of the lines and from their intensities the isotope shift is found

$$\delta\nu = 0{\cdot}0078 \text{ cm.}^{-1} \text{ for } S_{\frac{1}{2}}-{}^2P_{\frac{3}{2}},$$

$$\delta\nu = 0{\cdot}0074 \text{ cm.}^{-1} \text{ for } S_{\frac{1}{2}}-{}^2P_{\frac{1}{2}},$$

$$\text{Average} = 0{\cdot}0076 \pm 0{\cdot}0005 \text{ cm.}^{-1}.$$

The theoretical value of the centre of gravity shift, considering potassium as hydrogen-like, is 0·0087 cm.$^{-1}$, in the same direction as that observed.

The authors take this opportunity of thanking Professor Plaskett for placing at their disposal his excellent photometer, and Professor Lindemann for his continued interest in the research; and also Queen's College and St John's College for the stipends granted to one of them.

SUMMARY

1. The intensity ratio of the hyperfine structure components of the resonance lines of K 39 was measured by the method of absorption in an atomic beam. The value 1·45 found agrees with the value required by the spin $\frac{3}{2}$ if allowance is made for the overlapping by the lines of the 14 times rarer isotope 41. The component of longer wave-length was the stronger, showing that the nuclear magnetic moment of K 39 is positive.

2. By using three atomic beams in series, of collimation $\frac{1}{35}$, it was possible to resolve the lines of K 41 as two satellites on either side of the weak component of K 39. The ratio of the splittings of the lines of K 39 and K 41 is found to be 1·77, in good agreement with the value 1·80 found by Manley in magnetic deflexion experiments. The magnetic moment of K 41 is positive, like that of K 39.

The lines of K 41 have an isotope shift of $+0{\cdot}0076$ cm.$^{-1}$ relative to the lines of K 39. The theoretical centre of gravity shift, considering potassium as hydrogen-like, is $+0{\cdot}0087$ cm.$^{-1}$.

3. The Zeeman effect of the hyperfine structure of the line $4S_{\frac{1}{2}}-4\,^2P_{\frac{1}{2}}$ of K 39 was investigated. Each of the two π components was found to consist of four lines, proving that the nuclear spin of K 39 has the value $\frac{3}{2}$, in agreement with magnetic deflexion experiments and the measurements of the intensity ratio of the hyperfine structure lines. The observed positions of the lines are in close agreement with the positions required by the quantum theoretical formulae.

REFERENCES

Fox, M. and Rabi, F. 1935 *Phys. Rev.* **48**, 746–51.
Goudsmit, S. and Bacher, R. F. 1930 *Z. Phys.* **66**, 13–30.
Goudsmit, S. and Back, E. 1927 *Z Phys* **43**, 321
Jackson, D A and Kuhn, H 1935 *Proc Roy. Soc.* A, **148**, 335.
- 1936 *Nature, Lond*, **137**, 108
— — 1937*a* *Nature, Lond*, **140**, 276.
— — 1937*b* *Proc Roy Soc.* A, **158**, 372.
Manley, J H 1936 *Phys. Rev.* **49**, 921
Meissner, K. W. and Luft, K F. 1937 *Z Phys.* **106**, 362.
Millman, S. 1935 *Phys. Rev* **47**, 739.

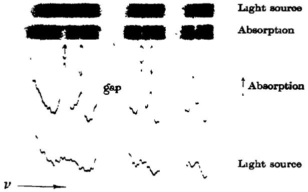


FIG. 1. 5 cm. etalon $4S_{\frac{1}{2}}-4\,{}^2P_{\frac{1}{2}}$.

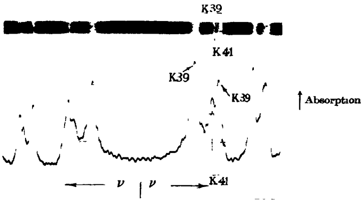


FIG. 2. 11 cm. etalon $4S_{\frac{1}{2}}-4\,{}^2P_{\frac{1}{2}}$. Triple atomic beam.

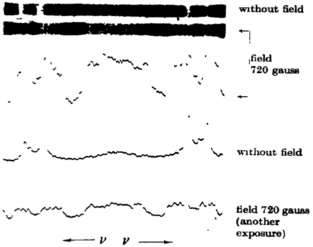


FIG. 3. 11 cm etalon $4S_{\frac{1}{2}}-4\,{}^2P_{\frac{1}{2}}$. Zeeman effect.

On the equations of electromagnetism

I. Identifications

By E. A. Milne, F R.S.

(*Received* 26 *July* 1937—*Abbreviated and revised* 29 *December* 1937)

1. The general object of the following papers is to ascertain what form the equations of electromagnetism take when derived on a purely kinematic basis. Maxwell's theory is not assumed. The only physical assumption made, namely, that a system of moving charges conserves its energy (defined kinematically) when the accelerations of the charges vanish, is a very slight one and is certainly satisfied in classical electromagnetism, but the resulting equations and laws, whilst coinciding with the classical theory to a considerable extent, differ in certain essential particulars. This arises from the avoidance of the empirical laws and hypothetical assumptions from which Maxwell's theory starts. In particular we avoid the formal inconsistency in the classical theory by which a magnetic intensity **H** is defined via the mechanical force on an isolated magnetic pole, yet isolated magnetic poles do not occur in the classical "theory of electrons". In the present treatment a magnetic intensity is defined via the mechanical force on a moving "charged" particle, as an element entering into the calculation of such force.

The general method is, adopting the dynamics constructed in previous papers on a purely kinematic basis (Milne 1936, 1937), to formulate equations of motion containing the next most general type of "external" force arising after "gravitational" forces have been dealt with. Such forces arise from the double differentiation of scalar "superpotentials", but we do not lay down what form these scalars are to take. Instead we allow them to determine themselves, by imposing the single physical assumption abovementioned, after the equation of energy has been derived. Once the scalar superpotentials have been so determined, their double differentiation yields symbols **E**, **H**, which are then compared with the empirical laws governing the interaction of "charges", this allows us to identify the adopted definition of charge and the symbols **E**, **H** with the similar quantities occurring in the experimental formulation. Lastly, we derive the identities satisfied by the resulting **E**, **H**; these partly coincide with, and partly differ from, the "field equations" with which the classical theory starts, and thus we end with theorems which play the part of the "laws of nature" assumed at the outset in the classical theory.

The scope of the work is that of the classical "theory of electrons", and the results should hold good down to phenomena for which the concept "particle" remains valid. The theory makes no attempt to deal with the ultimate structure of charges, to the analysis of which it is preliminary.

The analysis will be conducted in the first instance in t-measure, partly because the description of the universe is simplest in t-measure (hyperbolic space would be required if τ-dynamics were used), partly because relativistic invariance is readily ensured in this measure. We require our formulae to be valid for unrestricted velocities, and it involves no extra trouble to have them valid for unrestricted distances. The procedure justifies itself in the end, as the "field identities" replacing Maxwell's equations are found to hold good in t-measure without approximation. Purely mechanical relations, such as Coulomb's law, are found to take a simple form in τ-measure, but propagation equations are simplest in t-measure.

PROPERTIES OF 6-VECTORS

2. The 4-vectors at our disposal are constructed from the epoch t, position-vector $\mathbf{P}$ and velocity $\mathbf{V}$ of a particle relative to an arbitrary observer O situated on a particle of the substratum or smoothed-out universe. We proceed to construct force-vectors out of the 6-vectors which can be formed in conjunction with $\mathbf{P}$, t, $\mathbf{V}$. To recognize the scalar invariance of certain expressions which emerge it is convenient first to tabulate certain well-known properties of 6-vectors.

Transformations from one fundamental observer to another are included in the transformations for which

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = c^2dt^2 - dx^2 - dy^2 - dz^2 \quad (x^4 = ct,\ x^1 = x, \text{ etc.})$$

is invariant. With this ds^2, take a covariant skew-symmetric tensor $T_{\mu\nu}$, and write

$$\begin{matrix} T_{23} = H_x, & T_{31} = H_y, & T_{12} = H_z \\ T_{14} = E_x, & T_{24} = E_y, & T_{34} = E_z \end{matrix} \quad (T_{\mu\nu} = -T_{\nu\mu}).$$

Then $\mathbf{H}$, $\mathbf{E}$ are 3-vectors, and we call $T_{\mu\nu}$ the 6-vector $(\mathbf{H}, \mathbf{E})$. We associate with $T_{\mu\nu}$ the contravariant tensor $T^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}T_{\alpha\beta}$, the reciprocal covariant tensor $\bar{T}_{\mu\nu} = \frac{1}{2}A_{\mu\nu\alpha\beta}T^{\alpha\beta}$, and the reciprocal contravariant tensor $\bar{T}^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}\bar{T}_{\alpha\beta}$, A being the usual "alternate" tensor associated with ds^2.

If Q^α, R^α are contravariant 4-vectors, then their space-parts form

3-vectors $\mathbf{Q}, \mathbf{R}$; we write $Q^4 = Q_t$, $R^4 = R_t$. We can construct from Q^α and R^α the covariant skew-symmetric tensor $S_{\mu\nu} = A_{\mu\nu\alpha\beta}Q^\alpha R^\beta$ and its contravariant and reciprocal forms. Expressed in terms of 3-vectors and associated scalars, the complete system is given by the scheme

$$\begin{aligned}
T_{\mu\nu} &= (\mathbf{H}, \mathbf{E}), & S_{\mu\nu} &= (\mathbf{Q}R_t - \mathbf{R}Q_t, \mathbf{Q}\wedge\mathbf{R}),\\
T^{\mu\nu} &= (\mathbf{H}, -\mathbf{E}), & S^{\mu\nu} &= (\mathbf{Q}R_t - \mathbf{R}Q_t, -(\mathbf{Q}\wedge\mathbf{R})),\\
\tilde{T}_{\mu\nu} &= (-\mathbf{E}, \mathbf{H}), & \tilde{S}_{\mu\nu} &= (-(\mathbf{Q}\wedge\mathbf{R}), \mathbf{Q}R_t - \mathbf{R}Q_t),\\
\tilde{T}^{\mu\nu} &= (-\mathbf{E}, -\mathbf{H}), & \tilde{S}^{\mu\nu} &= (-(\mathbf{Q}\wedge\mathbf{R}), -(\mathbf{Q}R_t - \mathbf{R}Q_t)).
\end{aligned}$$

FORMATION OF FORCE-VECTORS

3. Put $Y = 1 - \mathbf{V}^2/c^2$. We generate contravariant force-vectors F^α from $(\mathbf{H}, \mathbf{E})$ and the available 4-vectors $(\mathbf{P}, ct)$ or say P^ν, and $(\mathbf{V}/Y^{\frac{1}{2}}, c/Y^{\frac{1}{2}})$ or say V^ν, by the relations

$$F^\alpha = g^{\alpha\mu}P^\nu T_{\mu\nu},$$

or

$$F^\alpha = g^{\alpha\mu}V^\nu T_{\mu\nu}.$$

The first yields the 4-vector

$$(ct\mathbf{E} + \mathbf{P}\wedge\mathbf{H},\ \mathbf{E}.\mathbf{P}), \tag{1}$$

the second the 4-vector

$$\left(c\frac{\mathbf{E}}{Y^{\frac{1}{2}}} + \frac{\mathbf{V}\wedge\mathbf{H}}{Y^{\frac{1}{2}}},\ \frac{\mathbf{E}.\mathbf{V}}{Y^{\frac{1}{2}}}\right). \tag{2}$$

Either of these may be multiplied by any scalar. After they have been multiplied by appropriate scalars, (2) must reduce to (1) when we put $\mathbf{V} = \mathbf{P}/t$. Thus (2) includes (1). It is therefore sufficient to consider (2). We multiply it by an arbitrary scalar k/c.

EQUATIONS OF MOTION

4. In the t-dynamics, the equations of motion of a particle of mass m at position P at epoch t moving with velocity $\mathbf{V}$ under an external force $(\mathbf{F}, F_t)$ are

$$\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}\left[m\xi^{\frac{1}{2}}\frac{\mathbf{V}}{Y^{\frac{1}{2}}}\right] = -\frac{m\xi^{\frac{1}{2}}}{X}\left(\mathbf{P} - \mathbf{V}\frac{Z}{Y}\right) + \mathbf{F}, \tag{3}$$

$$\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}\left[m\xi^{\frac{1}{2}}\frac{c}{Y^{\frac{1}{2}}}\right] = -\frac{m\xi^{\frac{1}{2}}}{X}\left(ct - c\frac{Z}{Y}\right) + F_t, \tag{3'}$$

where

$$X = t^2 - \mathbf{P}^2/c^2, \quad Y = 1 - \mathbf{V}^2/c^2, \quad Z = t - \mathbf{P}.\mathbf{V}/c^2, \tag{4}$$

$$\xi = Z^2/XY. \tag{5}$$

In (3) and (3′) the first term on each right-hand side represents the pull of the material of the universe, in t-measure, on the particle concerned; $m\xi^{\frac{1}{2}}$ is the inertial mass under the circumstances of motion.

Combination of (3) and (3′) yields the two energy-equations

$$\mathbf{F}.\left(\frac{\mathbf{V}}{Y^{\frac{1}{2}}}-\mathbf{P}\frac{Y^{\frac{1}{2}}}{Z}\right)-F_t\left(\frac{c}{Y^{\frac{1}{2}}}-ct\frac{Y^{\frac{1}{2}}}{Z}\right)=\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(mc^2\xi^{\frac{1}{2}}), \tag{6}$$

$$F_t\frac{c}{Y^{\frac{1}{2}}}-\mathbf{F}.\frac{\mathbf{V}}{Y^{\frac{1}{2}}}=\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(mc^2\xi^{\frac{1}{2}}). \tag{7}$$

If we take (2) multiplied by any scalar to represent a possible $(\mathbf{F}, F_t)$, relations (6) and (7) become self-inconsistent. Previous experience in finding forms for $\mathbf{F}$ suitable for representing gravitational forces suggests that (2) requires completing by the addition of a term arising from change of mass with velocity. We therefore examine the consequences of introducing as a form of external force the definitions

$$\mathbf{F}=\frac{k}{Y^{\frac{1}{2}}}\left[\mathbf{E}+\frac{\mathbf{V}\wedge\mathbf{H}}{c}\right]+\alpha\frac{\mathbf{V}}{Y^{\frac{1}{2}}}, \tag{8}$$

$$F_t=\frac{k}{Y^{\frac{1}{2}}}\frac{\mathbf{E}.\mathbf{V}}{c}+\alpha\frac{c}{Y^{\frac{1}{2}}}, \tag{8′}$$

where α is a scalar to be determined

ELECTROMAGNETIC FORCE-VECTOR AND ENERGY EQUATION

5. Introducing (8) and (8′) into (7) we find at once

$$\alpha=\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(m\xi^{\frac{1}{2}}). \tag{9}$$

Introducing (8) and (8′) into (6) we then find

$$\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(mc^2\xi^{\frac{1}{2}})=k\frac{Y^{\frac{1}{2}}}{Z}\left[\frac{\mathbf{E}\ (\mathbf{V}t-\mathbf{P})}{Y^{\frac{1}{2}}}-\frac{\mathbf{P}\wedge\mathbf{V}.\mathbf{H}}{cY^{\frac{1}{2}}}\right]. \tag{10}$$

Here the square bracket is an invariant, namely the double inner product $T_{\mu\nu}\bar{S}^{\mu\nu}$, where $S_{\mu\nu}$ is formed as above from the 4-vectors $(\mathbf{P}/c, t)$ and $(\mathbf{V}/Y^{\frac{1}{2}}, c/Y^{\frac{1}{2}})$.

It would be possible to retain the undetermined k until the time comes to choose it. We ultimately choose it so that a scalar invariant shall exist identifiable as the electromagnetic energy of a system of charges. Anticipating matters, without more ado we now choose

$$k = \frac{Z}{Y^{\frac{1}{2}}}\frac{e}{t_0}, \tag{11}$$

where e is a constant eventually identified as the "charge" on the moving particle P, and t_0 is a normalization constant of the dimensions of a time. When t_0 is chosen to be the present value of t, the number e will be found later to coincide with the measure of charge on the usual electrostatic system.

We now have for the external force $(\mathbf{F}, F_t)$ on a particle of charge e and mass m,

$$\mathbf{F} = \frac{e}{t_0}\frac{Z}{Y^{\frac{1}{2}}}\left[\mathbf{E} + \frac{\mathbf{V}\wedge\mathbf{H}}{c}\right]\frac{1}{Y^{\frac{1}{2}}} + \frac{\mathbf{V}}{Y^{\frac{1}{2}}}\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(m\xi^{\frac{1}{2}}), \tag{12}$$

$$F_t = \frac{e}{t_0}\frac{Z}{Y^{\frac{1}{2}}}\left[\frac{\mathbf{E}.\mathbf{V}}{c}\right]\frac{1}{Y^{\frac{1}{2}}} + \frac{c}{Y^{\frac{1}{2}}}\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(m\xi^{\frac{1}{2}}), \tag{12'}$$

and the energy equation (10) reduces to

$$\frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(mc^2\xi^{\frac{1}{2}}) = e\frac{t}{t_0}\frac{1}{Y^{\frac{1}{2}}}\left[\mathbf{E} + \frac{\mathbf{V}\wedge\mathbf{H}}{c}\right].\left(\mathbf{V} - \frac{\mathbf{P}}{t}\right). \tag{13}$$

When later identifications have been effected, (13) will represent the rate of increase of energy $d(mc^2\xi^{\frac{1}{2}})/dt$ as the work performed by the Larmor-Lorentz force $\mathbf{E} + (\mathbf{V}\wedge\mathbf{H})/c$ in pushing the particle with velocity $\mathbf{V} - \mathbf{P}/t$ relative to its immediate cosmic environment (which of course has the velocity $\mathbf{P}/t$). The secular factor t/t_0 should be noted.

In the above we have neglected the contribution to the force arising from the local gravitational field of the mass m. This is taken into account in the more general equations of §§ 19, 20, Part II.

FORMATION OF THE 6-VECTOR (H, E)

6. First-stage differentiations of scalars result only in the construction of external forces representing the local gravitational field of mass-distributions over and above the general gravitational field of the substratum. To obtain the next most general type of external force, take an undefined scalar ϕ_{21}, a function of two 4-vectors $(\mathbf{P}_1, ct_1)$, $(\mathbf{P}_2, ct_2)$ and of associated velocity vectors, and write

$$(\mathbf{H}_1)_x = \frac{\partial^2\phi_{21}}{\partial y_1\partial z_2} - \frac{\partial^2\phi_{21}}{\partial z_1\partial y_2}, \quad (\mathbf{E}_1)_x = \frac{1}{c}\left[\frac{\partial^2\phi_{21}}{\partial x_1\partial t_2} - \frac{\partial^2\phi_{21}}{\partial x_2\partial t_1}\right], \tag{14}$$

where $(\mathbf{H}_1, \mathbf{E}_1)$ is to be the value of $(\mathbf{H}, \mathbf{E})$ at the event $(\mathbf{P}_1, t_1)$ "due to" a "source" at the event $(\mathbf{P}_2, t_2)$; the scalar ϕ_{21} may be called a "superpotential". By its mode of construction $(\mathbf{H}_1, \mathbf{E}_1)$ is a 6-vector,* and we examine the consequences of taking it to represent the next most general type of external force after gravitational forces. Interchanging the suffixes 1 and 2 we obtain the 6-vector $(\mathbf{H}_2, \mathbf{E}_2)$ at the event $(\mathbf{P}_2, t_2)$ "due to" a "source" at the event $(\mathbf{P}_1, t_1)$, in terms of the associated superpotential ϕ_{12}.

MOTION OF A PAIR OF CHARGED PARTICLES

7. We now form the equations of motion of a charged particle $(m_1, e_1, \mathbf{P}_1, t_1, \mathbf{V}_1)$ in the presence of a charged particle $(m_2, e_2, \mathbf{P}_2, t_2, \mathbf{V}_2)$. Introducing (12), (12′) in (3), (3′) with appropriate suffixes we find, on cancelling a term,

$$\frac{m_1 \xi_1^{\frac{1}{2}}}{Y_1^{\frac{1}{2}}} \frac{d}{dt_1}\left[\frac{\mathbf{V}_1}{Y_1^{\frac{1}{2}}}\right] = -\frac{m_1 \xi_1^{\frac{1}{2}}}{X_1}\left(\mathbf{P}_1 - \mathbf{V}_1 \frac{Z_1}{Y_1}\right) + \frac{e_1}{t_0}\frac{Z_1}{Y_1^{\frac{1}{2}}}\left[\mathbf{E}_1 + \frac{\mathbf{V}_1 \wedge \mathbf{H}_1}{c}\right]\frac{1}{Y_1^{\frac{1}{2}}}, \tag{15}$$

$$\frac{m_1 \xi_1^{\frac{1}{2}}}{Y_1^{\frac{1}{2}}} \frac{d}{dt_1}\left[\frac{c}{Y_1^{\frac{1}{2}}}\right] = -\frac{m_1 \xi_1^{\frac{1}{2}}}{X_1}\left(ct_1 - c \frac{Z_1}{Y_1}\right) + \frac{e_1}{t_0}\frac{Z_1}{Y_1^{\frac{1}{2}}}\left[\frac{\mathbf{E}_1 . \mathbf{V}_1}{c}\right]\frac{1}{Y_1^{\frac{1}{2}}}. \tag{15′}$$

Of these, (15) reduces to (15′) on scalar multiplication by $\mathbf{V}_1/c$. If we multiply (15) scalarly by

$$\frac{\mathbf{V}_1}{Y_1^{\frac{1}{2}}} - \mathbf{P}_1 \frac{Y_1^{\frac{1}{2}}}{Z_1},$$

(15′) by

$$\frac{c}{Y_1^{\frac{1}{2}}} - ct_1 \frac{Y_1^{\frac{1}{2}}}{Z_1}$$

and subtract, we recover the energy-equation (13) with each symbol suffixed 1.

8. If in (13) we now substitute expressions (14) for $\mathbf{E}_1$ and $\mathbf{H}_1$, we obtain

$$\frac{d}{dt_1}(m_1 c^2 \xi_1^{\frac{1}{2}}) = \frac{e_1}{t_0}\left[\sum_{x,y,z} \frac{u_1 t_1 - x_1}{c}\left(\frac{\partial^2 \phi_{21}}{\partial x_1 \partial t_2} - \frac{\partial^2 \phi_{21}}{\partial x_2 \partial t_1}\right)\right.$$
$$\left. - \sum_{x,y,z} \frac{y_1 w_1 - z_1 v_1}{c}\left(\frac{\partial^2 \phi_{21}}{\partial y_1 \partial z_2} - \frac{\partial^2 \phi_{21}}{\partial z_1 \partial y_2}\right)\right], \tag{16}$$

where we have written (u_1, v_1, w_1) for $\mathbf{V}_1$.

---

* Any tensor may be expressed as the sum of a symmetric tensor and a skew-symmetric tensor. I have not succeeded in giving a physical interpretation to the symmetric tensor that can be derived by differentiation of ϕ_{21}.

For brevity we now introduce operators

$$L_{\mu\nu} \equiv t_\mu \frac{\partial}{\partial t_\nu} + \mathbf{P}_\mu \cdot \frac{\partial}{\partial \mathbf{P}_\nu}, \qquad (\mu, \nu = 1, 2).$$

Then (16) after some re-arrangement yields

$$\frac{d}{dt_1}(m_1 c^2 \xi_1^4) + \frac{e_1}{ct_0}\left(\frac{\partial}{\partial t_1} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_1}\right)(-L_{12}\phi_{21})$$
$$= -\frac{e_1}{ct_0}\left(\frac{\partial}{\partial t_2} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_2}\right)(\phi_{21} + L_{11}\phi_{21}). \qquad (17)$$

Writing down the corresponding relation for particle 2 by interchanging the suffixes 1 and 2 and adding we get

$$\frac{d}{dt_1}(m_1 c^2 \xi_1^4) + \frac{d}{dt_2}(m_2 c^2 \xi_2^4) + \left(\frac{\partial}{\partial t_1} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_1}\right)\left(-\frac{e_1}{ct_0} L_{12}\phi_{21}\right)$$
$$+ \left(\frac{\partial}{\partial t_2} + \mathbf{V}_2 \cdot \frac{\partial}{\partial \mathbf{P}_2}\right)\left(-\frac{e_2}{ct_0} L_{21}\phi_{12}\right)$$
$$= -\frac{e_1}{ct_0}\left(\frac{\partial}{\partial t_2} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_2}\right)(\phi_{21} + L_{11}\phi_{21}) - \frac{e_2}{ct_0}\left(\frac{\partial}{\partial t_1} + \mathbf{V}_2 \cdot \frac{\partial}{\partial \mathbf{P}_1}\right)(\phi_{12} + L_{22}\phi_{12}). \qquad (18)$$

By its mode of derivation this equation should be capable of expressing the rate of change of mechanical energy, $m_1 c^2 \xi_1^4 + m_2 c^2 \xi_2^4$, of the system of two particles. To give a meaning to "rate of change" when more than one particle are concerned, we must adopt some standard of simultaneity. Choose then $t_1 = t$, $t_2 = t$. Let us endeavour to choose the superpotentials ϕ_{21} and ϕ_{12} in such a way that

$$-\frac{e_1}{ct_0} L_{12}\phi_{21} = -\frac{e_2}{ct_0} L_{21}\phi_{12} = \Phi, \qquad (19)$$

say. Then since

$$\frac{d\Phi}{dt} = \frac{dt_1}{dt}\left[\frac{\partial}{\partial t_1} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_1} + \dot{\mathbf{V}}_1 \cdot \frac{\partial}{\partial \mathbf{V}_1}\right]\Phi + \frac{dt_2}{dt}\left[\frac{\partial}{\partial t_2} + \mathbf{V}_2 \cdot \frac{\partial}{\partial \mathbf{P}_2} + \dot{\mathbf{V}}_2 \cdot \frac{\partial}{\partial \mathbf{V}_2}\right]\Phi,$$

relation (18) may be rewritten as

$$\frac{d}{dt}[m_1 c^2 \xi_1^4 + m_2 c^2 \xi_2^4 + \Phi] = \left[\dot{\mathbf{V}}_1 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_1} + \dot{\mathbf{V}}_2 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_2}\right]$$
$$-\frac{e_1}{ct_0}\left(\frac{\partial}{\partial t_2} + \mathbf{V}_1 \cdot \frac{\partial}{\partial \mathbf{P}_2}\right)(\phi_{21} + L_{11}\phi_{21}) - \frac{e_2}{ct_0}\left(\frac{\partial}{\partial t_1} + \mathbf{V}_2 \cdot \frac{\partial}{\partial \mathbf{P}_1}\right)(\phi_{12} + L_{22}\phi_{12}). \qquad (20)$$

If Φ can be identified with electromagnetic energy, the left-hand side of (20) measures the rate of change of total energy, and the right-hand side consists of terms linear in the accelerations $\dot{\mathbf{V}}_1$ and $\dot{\mathbf{V}}_2$ together with terms independent of the accelerations. This suggests that we endeavour to choose ϕ_{21} and ϕ_{12} in such a way that the rate of change of total energy vanishes when the accelerations vanish. This is our one physical assumption, which is of course satisfied experimentally. We shall have succeeded in this if it is possible to choose ϕ_{21} and ϕ_{12} so that, in addition to (19),

$$\phi_{21} + L_{11}\phi_{21} \equiv 0, \tag{21}$$

$$\phi_{12} + L_{22}\phi_{12} \equiv 0, \tag{21'}$$

i e. so that ϕ_{21} is homogeneous and of degree -1 in the variables t_1, $\mathbf{P}_1$, and ϕ_{12} homogeneous and of degree -1 in the variables t_2, $\mathbf{P}_2$.

9. The superpotential ϕ_{21} is to determine the field at $(\mathbf{P}_1, t_1)$ due to e_2 at $(\mathbf{P}_2, t_2)$. It will therefore be expected to possess a singularity at $\mathbf{P}_1 = \mathbf{P}_2$, $t_1 = t_2$. Our previous experience of gravitational theory (Milne 1937*e*, formula (22)) at once suggests the scalar function

$$(X_{12}^2 - X_1 X_2)^{-\frac{1}{2}}, \tag{22}$$

where $\quad X_1 = t_1^2 - \mathbf{P}_1^2/c^2, \quad X_2 = t_2^2 - \mathbf{P}_2^2/c^2, \quad X_{12} = t_1 t_2 - \mathbf{P}_1 . \mathbf{P}_2/c^2. \quad (23)$

Since (22) is homogeneous and of degree -1 in $\mathbf{P}_1$, t_1 it will be a possible ϕ_{21} if multiplied by any scalar not containing $\mathbf{P}_1$, t_1. Similarly we may construct a possible ϕ_{12}.

Consider now (19). We find by direct differentiation that we have identically

$$L_{12}\frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} \equiv 0. \tag{24}$$

Hence $(X_{12}^2 - X_1 X_2)^{-\frac{1}{2}}$ behaves as a constant under the linear operator L_{12}. Hence if we put

$$\phi_{21} = e_2\frac{\psi_{21}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}, \quad \phi_{12} = e_1\frac{\psi_{12}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}, \tag{25}$$

where ψ_{21}, ψ_{12} are scalars, (19) will be satisfied if we can satisfy

$$L_{12}\psi_{21} = L_{21}\psi_{12}. \tag{26}$$

10. To find a simple solution of this identity, put

$$Z_1 = t_1 - \mathbf{P}_1 . \mathbf{V}_1/c^2, \quad Z_2 = t_2 - \mathbf{P}_2 . \mathbf{V}_2/c^2, \tag{27}$$

$$Z_{12} = t_1 - \mathbf{P}_1 . \mathbf{V}_2/c^2, \quad Z_{21} = t_2 - \mathbf{P}_2 . \mathbf{V}_1/c^2. \tag{27'}$$

Then $Z_1/Y_1^{\frac{1}{2}}$, $Z_2/Y_2^{\frac{1}{2}}$, $Z_{12}/Y_2^{\frac{1}{2}}$, $Z_{21}/Y_1^{\frac{1}{2}}$ are scalars. We find at once that

$$L_{12}Z_2 = Z_{12}, \quad L_{12}Z_{21} = Z_1,$$
$$L_{21}Z_1 = Z_{21}, \quad L_{21}Z_{12} = Z_2.$$

Hence $$L_{12}(Z_2Z_{21}) = Z_{12}Z_{21} + Z_1Z_2 = L_{21}(Z_1Z_{12}). \tag{28}$$

Accordingly we shall have a solution of (26) if we choose ψ_{21} proportional to $Z_2Z_{21}/Y_2^{\frac{1}{2}}Y_1^{\frac{1}{2}}$, and ψ_{12} proportional to $Z_1Z_{12}/Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}$.

11. We now choose definitely

$$\phi_{21} = -\tfrac{1}{2}e_2\frac{Z_2Z_{21}}{(X_{12}^2 - X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}, \tag{29}$$

$$\phi_{12} = -\tfrac{1}{2}e_1\frac{Z_1Z_{12}}{(X_{12}^2 - X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}, \tag{29'}$$

so that by (19) $$\Phi = \frac{1}{2}\frac{e_1e_2}{ct_0}\frac{Z_{12}Z_{21} + Z_1Z_2}{(X_{12}^2 - X_1X_2)^{\frac{1}{2}}Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}. \tag{30}$$

The energy-equation (18) now takes the form

$$\frac{d}{dt}(m_1c^2\xi_1^{\frac{1}{2}} + m_2c^2\xi_2^{\frac{1}{2}} + \Phi) = -\frac{dR}{dt}, \tag{31}$$

where dR/dt is the rate of loss of energy, i.e. the rate of radiation, given by

$$\frac{dR}{dt} = -\left[\dot{\mathbf{V}}_1\cdot\frac{\partial\Phi}{\partial\mathbf{V}_1} + \dot{\mathbf{V}}_2\cdot\frac{\partial\Phi}{\partial\mathbf{V}_2}\right]. \tag{32}$$

It is thus a physical consequence of the assumption that the rate of change of total energy shall vanish when the accelerations vanish, that the actual rate of change of total energy is linear in the accelerations. This result is in marked contrast to the classical theory, in which the rate of radiation is quadratic in the accelerations. It remains to see if the symbols e_1, e_2, Φ, and the **E** and **H** derived from ϕ_{21} and ϕ_{12} can be identified with charge-measures, energy and electric and magnetic intensities as used in ordinary physics. We have had practically no freedom of choice in the selection of the superpotentials ϕ_{21} and ϕ_{12}, and we have not arranged for them to reproduce the empirical laws of electromagnetic phenomena. We shall find how remarkably they do in fact reproduce them.

IDENTIFICATIONS

12. *Energy*. In virtue of the identity

$$(X_{12}^2 - X_1X_2)^{\frac{1}{2}} = \frac{1}{c}\left[(t_1\mathbf{P}_2 - t_2\mathbf{P}_1)^2 - \frac{(\mathbf{P}_1\wedge\mathbf{P}_2)^2}{c^2}\right]^{\frac{1}{2}},$$

we have for $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{P}_2| \ll ct_2$,

$$(X_{12}^2 - X_1 X_2)^{\frac{1}{2}} \sim c^{-1} |t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1|,$$

and also

$$Z_{12} \sim t_1, \quad Z_{21} \sim t_2, \quad Z_1 \sim t_1, \quad Z_2 \sim t_2.$$

Hence under these conditions and for $|\mathbf{V}_1| \ll c$, $|\mathbf{V}_2| \ll c$, we have from (30)

$$\Phi \sim e_1 e_2 \frac{t_1 t_2}{t_0} \frac{1}{|t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1|}.$$

For $t_1 = t_2 = t$, this gives

$$\Phi \sim \frac{t}{t_0} \frac{e_1 e_2}{|\mathbf{P}_2 - \mathbf{P}_1|}. \tag{33}$$

This should be the energy associated with a pair of charges e_1, e_2 at separation $|\mathbf{P}_2 - \mathbf{P}_1|$ in t-measure at time t. For $t = t_0$ it gives

$$\Phi \sim \frac{e_1 e_2}{|\mathbf{P}_2 - \mathbf{P}_1|}, \tag{34}$$

which is the standard formula of electrostatics for the energy associated with the two charged particles. This identifies our measures of charges e_1, e_2 as agreeing with the usual electrostatic units at the present epoch.

We now transform to τ-measure. If we regraduate all fundamental observers' clocks from t to τ according to the law

$$\frac{dt}{t} = \frac{d\tau}{t_0}, \quad \tau = t_0 \log (t/t_0) + t_0, \tag{35}$$

and adjust all derived measures accordingly, we know (Milne 1937*a*) that the τ-measure $|\mathbf{\Pi}_2 - \mathbf{\Pi}_1|$ of the separation $|\mathbf{P}_2 - \mathbf{P}_1|$ is given locally by

$$|\mathbf{P}_2 - \mathbf{P}_1| = \frac{t}{t_0} |\mathbf{\Pi}_2 - \mathbf{\Pi}_1|. \tag{35'}$$

But Φ, being an energy, is a time-invariant. Hence by (33)

$$\Phi \sim \frac{e_1 e_2}{|\mathbf{\Pi}_2 - \mathbf{\Pi}_1|}, \tag{36}$$

for any epoch τ Thus we derive the usual electrostatic formula as holding at all epochs τ in τ-measure.

Certain exact forms of the energy-formula are of interest. When the particles are at rest relative to the observer, $\mathbf{V}_1 = 0$ and $\mathbf{V}_2 = 0$, and $Z_{12} = Z_1 = t_1$, $Z_{21} = Z_2 = t_2$. Then (30) gives

$$\Phi = \frac{e_1 e_2}{t_0} \frac{t_1 t_2}{[(t_1 \mathbf{P}_2 - t_2 \mathbf{P}_1)^2 - c^{-2}(\mathbf{P}_1 \wedge \mathbf{P}_2)^2]^{\frac{1}{2}}}. \tag{37}$$

If we choose the observer O to be at one of the particles, say P_2, we have $\mathbf{P}_2 = 0$, and then

$$\Phi = \frac{e_1 e_2}{t_0} \frac{t_1}{|\mathbf{P}_1|} \tag{38}$$

independent of t_2. For $t_1 = t_0$ we have

$$\Phi = \frac{e_1 e_2}{|\mathbf{P}_1|}, \tag{39}$$

which is here an exact form of the empirical electrostatic formula.

In this exact formula, O is at P_2 and $\mathbf{V}_2 = 0$, consequently e_2 is at "local rest". But $\mathbf{V}_1 = 0$, not $\mathbf{V}_1 = \mathbf{P}_1/t_1$, and hence e_1 though at rest relative to O is not at cosmical "local rest". It follows that in electrostatics, "rest" has to be taken as meaning "rest relative to the observer", not "local rest". A distant charged particle at "local rest" is in motion relative to the observer, and so originates a magnetic field, as we shall see. We see that "rest" does not mean the same thing in dynamics and in electrostatics.

13. *Electric intensity*. It will be found as an unforeseen identity that

$$\left(\frac{\partial}{\partial \mathbf{P}_1}\frac{\partial}{\partial t_2} - \frac{\partial}{\partial \mathbf{P}_2}\frac{\partial}{\partial t_1}\right) \frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} \equiv 0, \tag{40}$$

whilst $$\frac{\partial}{\partial \mathbf{P}_1} Z_2 = 0, \quad \frac{\partial}{\partial \mathbf{P}_1} Z_{21} = 0, \quad \frac{\partial}{\partial t_1} Z_2 = 0, \quad \frac{\partial}{\partial t_1} Z_{21} = 0.$$

Making use of these relations, we find from (14) and (29) that the intensity E_1 at P_1 due to a charge e_2 at P_2 is given by

$$\begin{aligned}\mathbf{E}_1 &= \frac{1}{c}\left(\frac{\partial}{\partial \mathbf{P}_1}\frac{\partial}{\partial t_2} - \frac{\partial}{\partial \mathbf{P}_2}\frac{\partial}{\partial t_1}\right)\left[-\frac{1}{2}\frac{e_2}{Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}}\frac{Z_2 Z_{21}}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}}\right] \\ &= \frac{e_2}{c^3 Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{(\mathbf{P}_1 X_2 - \mathbf{P}_2 X_{12})\frac{1}{2}(Z_2 + Z_{21}) - (t_1 X_2 - t_2 X_{12})\frac{1}{2}(\mathbf{V}_1 Z_2 + \mathbf{V}_2 Z_{21})}{(X_{12}^2 - X_1 X_2)^{\frac{3}{2}}}.\end{aligned} \tag{41}$$

It will be noted that identity (40), which we did not arrange for beforehand, avoids the threatened appearance of too high a power in the denominator of (41), and so leads to the inverse square law, as we shall now see.

For $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{P}_2| \ll ct_2$, we have $t_1 X_2 - t_2 X_{12} \sim 0$, and if also $|\mathbf{V}_1| \ll c$, $|\mathbf{V}_2| \ll c$ the value of E_1 is given approximately by

$$\mathbf{E}_1 \sim e_2 \frac{t_2^2 (t_2 \mathbf{P}_1 - t_1 \mathbf{P}_2)}{|t_2 \mathbf{P}_1 - t_1 \mathbf{P}_2|^3}. \tag{42}$$

For $t_1 = t_2$, this reduces to

$$\mathbf{E}_1 \sim e_2 \frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|^3} \tag{43}$$

independent of time. This is as inverse square intensity directed from e_2 towards e_1, i.e. a repulsion, and so identifies our symbol $\mathbf{E}$ as in accordance with the empirical Coulomb law.

Formula (41) should give the exact relativistic form of the law. To see the meaning of the exact formula, choose O at P_2, so that $\mathbf{P}_2 = 0$; then $t_1 X_2 - t_2 X_{12} = 0$, $Z_2 = Z_{21} = t_2$ and $X_2 = t_2^2$. Accordingly (41) becomes

$$\mathbf{E}_1 = \frac{e_2}{Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}} \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}, \tag{44}$$

the inverse square law with a velocity factor depending on both source-velocity $\mathbf{V}_2$ and the velocity $\mathbf{V}_1$ of the test-charge e_1.

It is to be noted that the exact relativistic form of the inverse square law for electrostatics is different from that for gravitation (Milne 1937*e*, p. 9). It may be mentioned further that whilst inertial masses are essentially of the same sign, and so (as shown in previous papers) gravitational forces are essentially attractive, there is no restriction on the sign of charges e, and so both attraction and repulsion are permissible.

14. *Mechanical force due to electric intensity.* In the t-description, the measure of intensity must be sharply distinguished from the measure of the mechanical force to which it gives rise. By (12) this mechanical force has a component contributed by $\mathbf{E}_1$ of amount

$$\mathbf{F}_1 = \frac{e_1}{t_0} \frac{Z_1}{Y_1^{\frac{1}{2}}} \frac{\mathbf{E}_1}{Y_1^{\frac{1}{2}}}. \tag{45}$$

For $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{V}_1| \ll c$, this is approximately

$$\mathbf{F}_1 \sim e_1 \frac{t_1}{t_0} \mathbf{E}_1. \tag{46}$$

Hence by (44) the mechanical force on e_1 due to a charge e_2 at rest at the origin is given by

$$\mathbf{F}_1 \sim e_1 e_2 \frac{t_1}{t_0} \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}, \tag{47}$$

in which the secular factor (t_1/t_0) should be noted.

Let $\mathbf{\Phi}_1$ be the τ-measure of $\mathbf{F}_1$. Then by the general dynamical theory (Milne 1937*a*, p. 338) for local phenomena

$$\mathbf{\Phi}_1 \sim \frac{t_1}{t_0}\mathbf{F}_1. \tag{48}$$

Accordingly, using (35′),

$$\mathbf{\Phi}_1 \sim e_1 e_2 \left(\frac{t_1}{t_0}\right)^2 \frac{\mathbf{P}_1}{|\mathbf{P}_1|^3} = e_1 e_2 \frac{\mathbf{\Pi}_1}{|\mathbf{\Pi}_1|^3}. \tag{49}$$

Thus in τ-measure the secular factor disappears, and we get the usual inverse square expression for $\mathbf{\Phi}_1$. This shows that as far as electric intensities are concerned, the empirical laws relating to mechanical effects employ τ-measure.

15. *Magnetic intensity.* It will be found that identically

$$\left(\frac{\partial}{\partial y_1}\frac{\partial}{\partial z_2} - \frac{\partial}{\partial y_2}\frac{\partial}{\partial z_1}\right)\frac{1}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}}} = 0, \tag{50}$$

whilst $\quad \dfrac{\partial}{\partial y_1}Z_2 = 0, \quad \dfrac{\partial}{\partial z_1}Z_2 = 0, \quad \dfrac{\partial}{\partial y_1}Z_{21} = 0, \quad \dfrac{\partial}{\partial z_1}Z_{21} = 0.$

Hence by (14) and (29), after some straightforward differentiation it is found that

$$\mathbf{H}_1 = \frac{e_2}{c^4}\frac{1}{Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}\frac{\frac{1}{2}(Z_{21}\mathbf{V}_2 + Z_2\mathbf{V}_1)\wedge(X_2\mathbf{P}_1 - X_{12}\mathbf{P}_2)}{(X_{12}^2 - X_1 X_2)^{\frac{3}{2}}}. \tag{51}$$

This, the complementary formula to (41), should give in relativistic form the magnetic intensity $\mathbf{H}_1$ due to a charge e_2 at P_2 moving with velocity $\mathbf{V}_2$, as measured by a test-charge e_1 at P_1 moving with velocity $\mathbf{V}_1$. It will be found to give new features in the description of a magnetic field.

Consider the form which (51) takes for local phenomena and small velocities. For $|\mathbf{P}_1| \ll ct_1$, $|\mathbf{P}_2| \ll ct_2$, $|\mathbf{V}_1| \ll c$, $|\mathbf{V}_2| \ll c$, it gives

$$\mathbf{H}_1 \sim \frac{e_2}{c}\frac{\frac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2)\wedge t_2^2(t_2\mathbf{P}_1 - t_1\mathbf{P}_2)}{|t_1\mathbf{P}_2 - t_2\mathbf{P}_1|^3}, \tag{52}$$

or, for $t_1 = t_2$,

$$\mathbf{H}_1 \sim \frac{e_2}{c}\tfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2)\wedge\frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|^3}. \tag{53}$$

The exact form of (51) for $\mathbf{P}_2 = 0$ is found to be

$$\mathbf{H}_1 = \frac{e_2}{c}\frac{1}{Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}}}\tfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2)\wedge\frac{\mathbf{P}_1}{|\mathbf{P}_1|^3}. \tag{54}$$

Both (53) and (54) differ from the classical form of the law giving the magnetic field due to a moving charge, usually called the "law of Biot and Savart". Biot and Savart's law states that, apart from relativistic refinements and corrections due to "retardation" effects, the magnetic intensity due to a charge e_2 at P_2 moving with velocity $\mathbf{V}_2$ relative to the observer, as evaluated at position P_1, is given by

$$H_1 = \frac{e_2}{c}\mathbf{V}_2 \wedge \frac{\mathbf{P}_1 - \mathbf{P}_2}{|\mathbf{P}_1 - \mathbf{P}_2|^3} \tag{55}$$

Our formulae (53) and (54) differ in having $\frac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2)$ in place of $\mathbf{V}_2$. In the classical empirical law, the magnetic field of a moving charge depends on the velocity of that charge but not on the velocity of the test-charge used to measure the field; in the present deductive treatment, the magnetic field of a moving charge depends both on the velocity of the source-charge and on that of the test-charge.

16. In the classical theory, a magnetic field is measured by the mechanical force on an isolated magnetic pole at the point considered. But isolated magnetic poles do not exist, and in a fundamental account of electromagnetism we must aim at proceeding without introducing them. This we have done—our deductive procedure has thrown up no entity which could be taken to represent an isolated magnetic pole In our work, a magnetic field first appears as a constituent of the external force acting on a moving test-charge, and (53) and (54) may be taken to be the consequences.

17. Consider some examples of (53) If we take the test-charge at rest relative to the moving charge, we have $\mathbf{V}_1 = \mathbf{V}_2$, and then (53) gives just the classical formula. But if $\mathbf{V}_1 \neq \mathbf{V}_2$, writing

$$\tfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2) = \mathbf{V}_2 + \tfrac{1}{2}(\mathbf{V}_1 - \mathbf{V}_2), \tag{56}$$

we see that (53) gives the classical contribution together with a contribution proportional to the relative velocity of the two charges and equal numerically to just one-half of what would be calculated by a crude application of the classical formula to the relative velocity.

The physical interpretation of these results is clear. A magnetic field is originated by charges in motion. If a test-charge has a velocity unequal to that of a given source-charge, the relative motion of the two will originate a magnetic field at the test-charge, which will contribute a constituent to the mechanical force acting on the test-charge. Thus the magnetic field due to a given source in given motion, at some given point, as reckoned by a given observer, may be resolved into two parts, one due to the motion of the source relative to the observer (this is the classical contribution), and one due to the relative velocity of test-charge and source-charge.

This discussion makes it doubtful whether the classical formula (55) correctly takes into account the *two* relative motions concerned, namely, the relative motion of observer and source and the relative motion of observer and test-charge; (55) appears only to be correct for zero relative motion of test-charge and source-charge. I suggest that (55) should be replaced by (53). I have considered carefully whether there is any possible modification of the analysis determining the superpotentials which would lead to a result in accordance with the usual Biot and Savart law, but I can find no escape from the present conclusion. It may be noted that (51) and (41) display $(\mathbf{H}_1, \mathbf{E}_1)$ as of the form $\mathcal{S}_{\mu\nu}$, (§ 2), with

$$\mathbf{Q} = -(Z_{21}\mathbf{V}_2 + Z_2\mathbf{V}_1)/Y_1^{\frac{1}{2}}Y_2^{\frac{1}{2}} \text{ and } \mathbf{R} = X_2\mathbf{P}_1 - X_{12}\mathbf{P}_2.$$

The usual form of the Biot and Savart law is scarcely capable of being brought into 6-vector parallelism with the inverse square form for E_1.

18. If, now, the intensity of a magnetic field $\mathbf{H}_1$ at P_1 depends on the velocity $\mathbf{V}_1$ of the test-charge used to measure it, how comes it about that physicists can in fact find definite measures for magnetic fields? To answer this, consider (53) more closely. The magnetic intensity $\mathbf{H}_1$ at P_1 due to a system of moving charges $e_2, \ldots, e_s, \ldots, e_n$ at $P_1, \ldots, P_s, \ldots, P_n$ moving with velocities $\mathbf{V}_2, \ldots, \mathbf{V}_s, \ldots, \mathbf{V}_n$ will be given, neglecting relativistic refinements, by

$$\mathbf{H}_1 = \sum_{s=2}^{n} \frac{e_n}{c} \tfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_s) \wedge \frac{\mathbf{P}_1 - \mathbf{P}_s}{|\mathbf{P}_1 - \mathbf{P}_s|^3}, \tag{57}$$

whilst the electric intensity $\mathbf{E}_1$ will be given, by (43), by

$$\mathbf{E}_1 = \sum_{s=2}^{n} e_s \frac{\mathbf{P}_1 - \mathbf{P}_s}{|\mathbf{P}_1 - \mathbf{P}_s|^3}. \tag{57'}$$

Hence

$$\mathbf{H}_1 = \tfrac{1}{2} \sum_{s=2}^{n} \frac{e_s}{c} \mathbf{V}_s \wedge \frac{\mathbf{P}_1 - \mathbf{P}_s}{|\mathbf{P}_1 - \mathbf{P}_s|^3} + \frac{1}{2} \frac{\mathbf{V}_1 \wedge \mathbf{E}_1}{c}. \tag{58}$$

In this, the first term is just one-half what would be calculated for the magnetic intensity from a direct application of the empirical Biot and Savart law. We may call it $\frac{1}{2}(\mathbf{H}_1)_{\text{class}}$.

Now the electromagnetic systems whose external magnetic fields are measured by physicists usually have the property that they behave at external points as if *electrostatically neutral*. Examples of such systems are afforded by permanent magnets and conductors carrying currents. For such systems we have $\mathbf{E}_1 = 0$. Relation (58) then becomes

$$\mathbf{H}_1 = \frac{1}{2}\sum_{s=2}^{n} \frac{e_s}{c}\mathbf{V}_s \wedge \frac{\mathbf{P}_1-\mathbf{P}_s}{|\mathbf{P}_1-\mathbf{P}_s|^3} = \frac{1}{2}(\mathbf{H}_1)_{\text{class}}. \tag{59}$$

We see that for such systems there is an objective value of $\mathbf{H}_1$ independent of the velocity $\mathbf{V}_1$ of the test-charge e_1 used to measure the field. Hence the field can be described, and its value stated, without mention of $\mathbf{V}_1$: the magnetic field in such cases possesses a determinate intensity.

The apparent outstanding exception is Rowland's classical experiment (1876, 1889) on the magnetic field produced by rotating a charged disk. This is usually taken to demonstrate the validity of Biot and Savart's law as applied to a charged system. But it must be remembered that Rowland measured the field by means of a small magnet. By (59), such a small magnet must on our treatment be credited with twice the number of elementary circulating charges it would have on the classical theory. This cuts out our factor $\frac{1}{2}$ arising in our treatment of the field of the revolving disk-charges, and so Rowland would get agreement with the Biot and Savart law. (For a small magnet, the mean value of $\mathbf{V}_1$ in (57) is zero, hence (57) gives one-half the classically calculated value.)

It will be clear that as long as we use test-magnets to measure fields (magnets whose moments have been measured by their mechanical effects) our formulation of the Biot and Savart law makes no difference in the resulting values of $\mathbf{H}_1$, for electrostatically neutral systems. A difference comes in only when we calculate backwards to the charges originating the field (or to their velocities). It is clear that we must look for experimental discrepancies with the classical theory only in the interiors of systems of moving charges, or when the field-measuring apparatus is an isolated charge. Such systems are found producing the magnetic fields in the interiors of atomic or molecular systems, where the difficulties met with in applying classical electromagnetism are notorious.

19. In the simple valence-electron model of an atom, the series-electron is moving in the field of the core. The core may be considered to be a charge at rest relative to the observer, and the series-electron may be considered to be the test-charge. Applying formula (53) with $\mathbf{V}_2 = 0$, $\mathbf{V}_1 \neq 0$, we get

$$\mathbf{H}_1 = \frac{1}{2}\frac{e_2}{c}\mathbf{V}_1 \wedge \frac{\mathbf{P}_1-\mathbf{P}_2}{|\mathbf{P}_1-\mathbf{P}_2|^3}. \tag{60}$$

This again is just one-half what would be calculated classically by regarding the series-electron as at rest and the core as moving with relative velocity

$\mathbf{V}_1$ and so generating a magnetic field at the electron. In 1926 Uhlenbeck and Goudschmidt in this way applied the classical formula for the magnetic field at the electron due to the motion relative to the core, and obtained doublet separations for the corresponding lines *twice* those observed. This is in accordance with (60), and may be taken as experimental evidence in favour of (53). The factor $\frac{1}{2}$ was explained by L. H. Thomas (1926, 1927) in a somewhat complicated paper on the kinematics of an electron with an axis, as arising from Lorentz transformations to the successive frames defined by the successive positions of the moving electron. Thomas's explanation, like ours, arises from the fundamental embodiment of Lorentz transformations in the analysis, but we have not found it necessary to assign an axis to the electron. Our explanation has arisen in an unforced and entirely unforeseen way, and does not turn on the particular properties of the electrodynamic system considered; instead, it turns on a proper treatment of the relativity of magnetic fields combined with a proper epistemological consideration of how such fields can be known to an observer who has only charges and not magnets at his disposal.

Since 1927, the effect has been supposed to be explained by the attribution of *spin* to an electron. If electron spin is considered as an *ad hoc* assumption introduced to explain certain experimental facts not otherwise explained by classical electromagnetism, then it has no place in our rational kinematic treatment, which is concerned only with the consequences of a simple energy-assumption, and we have seen that certain phenomena at least can be understood without it, if other phenomena are found still requiring it, it must be formulated in such a way that its introduction does not disturb explanations naturally arising in the treatment which does not make use of it. If on the other hand electron spin is considered as a rational deduction from classical electrodynamics, it could equally well be deduced in some appropriately modified form from our rational reformulation of electrodynamics. This is beyond the scope of the present paper. It may be remarked that in the Zeeman effect, for which electron spin was largely invented, each electron's spin makes a contribution; in our form of the Biot and Savart formula each electron, possessing its own velocity $\mathbf{V}_1$, contributes to the magnetic field in which it finds itself. The manifest difficulties of the concept of electron spin, with its assignment of *structure* to an entity whose detailed equilibrium is not further considered, need not here be enlarged on.

20. The factor $\frac{1}{2}$ turns up in other magnetic phenomena. For example (Stoner 1930) experiments on the gyromagnetic effect suggest that the magnetization of ferromagnetics is entirely due to the "intrinsic spin"

of the electrons, and that their orbital moment is not effective. The present investigation suggests that the contribution of the orbital motion to the magnetic moment is one-half that calculated classically from the usual formula, and that the hypothesis of electron spin is unnecessary; this may be seen by noting that for an oscillatory path near a fixed point P_1, the time-mean of $\mathbf{V}_1$ will be zero, and so the mean value of $\frac{1}{2}(\mathbf{V}_1+\mathbf{V}_2)$ will be just $\frac{1}{2}\mathbf{V}_2$, which is effectively the same as saying that the magnetic moment will appear halved.

21. Formula (53) leads to a fundamental change in the concept of a magnetic field at a point. In general, not only does a magnetic field only exist when there is a test-charge at the point to measure it, but the value of the magnetic intensity depends on the velocity of the test-charge relative to the observer, the exception being when the given system of moving charges behaves as if electrostatically neutral, in which case the intensity is one-half that calculated classically. It may be expected in consequence that the field-equations of electromagnetism may require modification. We shall determine the "field-identities" satisfied by our $\mathbf{E}_1$ and $\mathbf{H}_1$ in Part II; the remarkable circumstance will emerge that though the value of $\mathbf{H}_1$ depends on $\mathbf{V}_1$, the *forms* of these field identities do not involve $\mathbf{V}_1$ explicitly. An account of the relations satisfied by $\mathbf{E}_1$ and $\mathbf{H}_1$ is thus possible whatever velocity $\mathbf{V}_1$ is assigned to the test-charge. In the case of the field arising from a single moving charge e_2 alone, we could always choose $\mathbf{V}_1 = \mathbf{V}_2$ and so get the classical value of $\mathbf{H}_1$; but when the system consists of a set of charges $e_2, \ldots, e_n$, no such choice is possible, and the general formula is necessary. The electrostatic intensity $\mathbf{E}_1$ depends normally on $\mathbf{V}_1$ just as $\mathbf{H}_1$ does, but the term in $\mathbf{V}_1$ is numerically insignificant.

It may be remarked that the emergence of (43) and (53) from the definition (14) of $\mathbf{E}_1$ and $\mathbf{H}_1$ affords a delicate test of the appropriateness of our choices of ϕ_{21} and ϕ_{12}, since the value of $\mathbf{H}_1$, for example, arises entirely from the terms in $1/c^2$ in the numerator of ϕ_{21}.

22. *Mechanical force due to magnetic field.* From (12), the component of mechanical force due to a magnetic field $\mathbf{H}_1$ for $|\mathbf{P}_1| \ll ct_1$ and $|\mathbf{V}_1| \ll c$ reduces to

$$\mathbf{F}_1 \sim e_1 \frac{t_1}{t_0} \frac{\mathbf{V}_1 \wedge \mathbf{H}_1}{c}, \tag{61}$$

a secular factor t_1/t_0 being required in t-measure as in the component due to the electric field. The mechanical force due to the magnetic effect of a charge e_2 at P_2 moving with velocity $\mathbf{V}_2$ is then approximately

$$\mathbf{F}_1 \sim \frac{e_1 e_2}{c^2}\frac{t_1}{t_0}\mathbf{V}_1 \wedge \frac{\frac{1}{2}(\mathbf{V}_1+\mathbf{V}_2)\wedge(\mathbf{P}_1-\mathbf{P}_2)}{|\mathbf{P}_1-\mathbf{P}_2|^3}. \tag{62}$$

The secular factor disappears in τ-measure.

This completes our identification of the symbols e, as representing charge, Φ as representing energy, $\mathbf{E}_1$ and $\mathbf{H}_1$ as representing electric and magnetic intensities, and the corresponding expressions giving the mechanical forces. The factor $\frac{1}{2}$ in our choice of ϕ_{21} and ϕ_{12} was of course made to make the unit of e come right in the energy formula Φ; the subsequent appropriateness of the other formulae is then not under our control. The revised formula for a magnetic field may be considered as the result of a mathematical experiment in the deductive method.

SUMMARY

The paper offers a purely kinematic formulation of the phenomena of electromagnetism. It proceeds by first examining the most general type of external forces after gravitational forces which are encountered in the study of the dynamics of a particle in the presence of the substratum or smoothed-out universe. Such external forces involve mention of 6-vectors, and the form of the resulting force is obtained in a form holding good for all epochs and for all distances from the observer. For the present epoch and at small distances it reduces to the Larmor-Lorentz formula for the ponderomotive force on a moving charge. A scalar multiplier emerges in the treatment which is afterwards identified as charge. A 6-vector is then derived by appropriate differentiation of an undetermined scalar with regard to the co-ordinates of two events, one being the source of the field and the other the event of the measurement of the field. With this it is possible to obtain the relativistic equations of motion of two charged particles in one another's presence, and to deduce the associated energy equation. By requiring the radiation to vanish when the accelerations vanish, it is found possible to fix all undetermined scalars completely. It is then possible to evaluate the "field" due to a given charge in given motion, at a given external point, as measured by the force on a test-charge in given motion, and by comparing the theoretically derived formulae with the observed experimental laws to identify the various symbols introduced. The inverse square law of Coulomb is obtained in exact relativistic form. But the law expressing the magnetic field at a

distance r from a given charge e_2 moving with a given velocity $\mathbf{V}_2$ relative to a fundamental observer O is found to contain mention of the velocity $\mathbf{V}_1$ of the test-charge measuring the field; in detail, the classical Biot and Savart law $\mathbf{H} = (e_2/c)\,\mathbf{V}_2 \wedge \mathbf{r}/|\,\mathbf{r}\,|^3$ is found to be replaced by

$$\mathbf{H} = (e_2/c)\,\tfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2) \wedge \mathbf{r}/|\,\mathbf{r}\,|^3.$$

The two agree for $\mathbf{V}_1 = \mathbf{V}_2$, but the new formula gives half the classically calculated field for $\mathbf{V}_2 = 0$, as required by Uhlenbeck and Goudschmidt and as found by a different method by L. H. Thomas.

REFERENCES

Milne 1936 *Proc. Roy. Soc.* A, **154**, 22.
— 1937*a* *Proc. Roy. Soc.* A, **158**, 324.
— 1937*b* *Proc. Roy. Soc.* A, **159**, 171.
— 1937*c* *Proc. Roy. Soc.* A, **159**, 526.
— 1937*d* *Quart. J. Math.* **8**, 22.
— 1937*e* *Proc. Roy. Soc.* A, **160**, 1.
Rowland 1876 *Ann. Phys., Lpz.*, **158**, 87.
Rowland and Hutchinson 1889 *Phil. Mag.* **5**, 27, 445.
Stoner 1930 "Magnetism", p. 68. London.
Thomas 1926 *Nature, Lond.*, **117**, 514.
1927 *Phil. Mag.* **7**, 3, 1.
Uhlenbeck and Goudschmidt 1926 *Nature, Lond*, **117**, 264.

On the equations of electromagnetism
II. Field Theory

By E. A. Milne, F.R.S.

(*Received* 26 *July* 1937—*Abbreviated and revised* 29 *December* 1937)

1. The object of the following paper is to enquire what general relations are satisfied by $\mathbf{E}_1$, $\mathbf{H}_1$, the electric and magnetic intensities as measured by a test-charge at P_1 at epoch t_1 moving with velocity $\mathbf{V}_1$, all relative to a given observer O, when it is in the presence of a system of moving charges $e_s\,(2 \leqslant s \leqslant n)$ which at events $(\mathbf{P}_s, t_s)$ have velocities $\mathbf{V}_s$

2. We begin by generalizing formulae previously obtained for a system of two charges moving in one another's presence. The electric and magnetic intensities $\mathbf{E}_1$, $\mathbf{H}_1$ at P_1 are given by

$$(\mathbf{E}_1)_x = \sum_{s=2}^{n} \frac{1}{c}\left[\frac{\partial^2 \phi_{s1}}{\partial x_1 \partial t_s} - \frac{\partial^2 \phi_{s1}}{\partial x_s \partial t_1}\right], \tag{1}$$

$$(\mathbf{H}_1)_x = \sum_{s=2}^{n} \left[\frac{\partial^2 \phi_{s1}}{\partial y_1 \partial z_s} - \frac{\partial^2 \phi_{s1}}{\partial z_1 \partial y_s}\right], \tag{1'}$$

where

$$\phi_{s1} = -\frac{1}{2}\frac{e_s}{Y_1^{\frac{1}{2}} Y_s^{\frac{1}{2}}} \frac{Z_s Z_{s1}}{(X_{1s}^2 - X_1 X_s)^{\frac{1}{2}}} \tag{2}$$

and

$$X_1 = t_1^2 - \mathbf{P}_1^2/c^2, \quad X_s = t_s^2 - \mathbf{P}_s^2/c^2, \quad X_{1s} = t_1 t_s - \mathbf{P}_1 . \mathbf{P}_s/c^2, \tag{3}$$

$$Y_1 = 1 - \mathbf{V}_1^2/c^2, \quad Y_s = 1 - \mathbf{V}_s^2/c^2, \tag{4}$$

$$Z_s = t_s - \mathbf{P}_s . \mathbf{V}_s/c^2, \quad Z_{s1} = t_s - \mathbf{P}_s . \mathbf{V}_1/c^2. \tag{5}$$

The field $(\mathbf{H}_s, \mathbf{E}_s)$ at any other charge of the system is to be calculated similarly from all the other charges of the system *together with the test-charge* e_1 at $(\mathbf{P}_1, t_1)$ moving with $\mathbf{V}_1$. It is an essential feature of our treatment that the test-charge is considered as an intrinsic part of the whole system. The events $(\mathbf{P}_s, t_s)$ are formally independent. We shall later find the relations between the t_s's required to evaluate the field in any actual case, but it is a feature of the "field identities" about to be obtained that they do not depend on imposing any particular relations between the t_s's.

The mechanical forces on the charged particles are to be obtained by equations of the type (12), (12') of Part I, which contain the normalization constant t_0. The equations of motion of each particle, of mass m_s, are then given by equations of the form (3), (3′) of Part I. Forming the associated energy equations and reducing them as in Part I, we may combine them to yield the relation

$$\frac{d}{dt}\left[\sum_{s=1}^{n} m_s c^2 \xi_s^{\frac{1}{2}} + \Phi\right] = \sum_{s=1}^{n} \dot{\mathbf{V}}_s \cdot \frac{\partial \Phi}{\partial \mathbf{V}_s}, \tag{6}$$

where

$$\xi_s = Z_s^2/X_s Y_s, \tag{7}$$

$$\Phi = \sum_{r,s} \Phi_{r,s}, \quad (r \neq s) \tag{8}$$

and

$$\Phi_{r,s} = \frac{1}{2}\frac{e_r e_s}{c t_0}\frac{Z_{rs} Z_{sr} + Z_r Z_s}{(X_{rs}^2 - X_r X_s)^{\frac{1}{2}} Y_r^{\frac{1}{2}} Y_s^{\frac{1}{2}}}. \tag{9}$$

In equation (6) it is to be understood that dt_s/dt is put equal to unity after the differentiations have been performed. Φ is to be interpreted as electromagnetic energy.

FIELD IDENTITIES

3. We shall use the notations div, curl in the senses

$$\operatorname{div} \mathbf{H}_1 = \frac{\partial(\mathbf{H}_1)_x}{\partial x_1} + \frac{\partial(\mathbf{H}_1)_y}{\partial y_1} + \frac{\partial(\mathbf{H}_1)_z}{\partial z_1},$$

$$\operatorname{curl} \mathbf{H}_1 = \frac{\partial(\mathbf{H}_1)_z}{\partial y_1} - \frac{\partial(\mathbf{H}_1)_y}{\partial z_1},$$

the differentiations being with respect to the space-co-ordinates x_1, y_1, z_1 of the event at which $\mathbf{E}_1$, $\mathbf{H}_1$ are evaluated. We use grad similarly.*

By (1) and (1′) we have

$$\begin{aligned}\operatorname{div} \mathbf{H}_1 &= \sum_{x,y,z} \sum_{s=2}^{n} \frac{\partial}{\partial x_1}\left[\frac{\partial^2 \phi_{s1}}{\partial y_1 \partial z_s} - \frac{\partial^2 \phi_{s1}}{\partial z_1 \partial y_s}\right], \\ &\equiv 0,\end{aligned} \tag{10}$$

and

$$\begin{aligned}(\operatorname{curl} \mathbf{E}_1)_x &= \frac{1}{c}\frac{\partial}{\partial y_1}\left[\sum_{s=2}^{n}\left(\frac{\partial^2 \phi_{s1}}{\partial z_1 \partial t_s} - \frac{\partial^2 \phi_{s1}}{\partial z_s \partial t_1}\right)\right] - \frac{1}{c}\frac{\partial}{\partial z_1}\left[\sum_{s=2}^{n}\left(\frac{\partial^2 \phi_{s1}}{\partial y_1 \partial t_s} - \frac{\partial^2 \phi_{s1}}{\partial y_s \partial t_1}\right)\right] \\ &= -\frac{1}{c}\frac{\partial}{\partial t_1}\left[\sum_{s=2}^{n}\left(\frac{\partial^2 \phi_{s1}}{\partial y_1 \partial z_s} - \frac{\partial^2 \phi_{s1}}{\partial z_1 \partial y_s}\right)\right] \\ &= -\frac{1}{c}\frac{\partial(\mathbf{H}_1)_x}{\partial t_1}.\end{aligned} \tag{11}$$

---

* Strictly speaking we should write div_1, curl_1, grad_1, but the suffix 1 may be understood.

4. Identities (10) and (11) contain no mention of the charges $e_2, \ldots, e_n$ giving rise to the field, or of their positions, epochs or velocities, or of the velocity $\mathbf{V}_1$ of the test-charge e_1, although $\mathbf{E}_1$, $\mathbf{H}_1$ depend on $\mathbf{V}_1$ as well as on the sources. We can if we like now drop the suffix 1, and so we arrive at identities which are identical with two of Maxwell's equations. The satisfaction of (10) and (11) does not depend on the structure of the ϕ_{s1}'s but only on the general tensor forms of $\mathbf{E}_1$ and $\mathbf{H}_1$.

5. In our formulation, relations (10) and (11) are identities, whereas in Maxwell's formulation they are empirical laws determined as inductions from experience. The Maxwell equation corresponding to (10) expresses the non-existence of isolated magnetic poles, the Maxwell equation corresponding to (11) expresses Faraday's law of electromagnetic induction.

6. We proceed to examine whether Maxwell's other two equations are satisfied in our formulation. We have by (1)

$$\begin{aligned}\operatorname{div}\mathbf{E}_1 &= \frac{1}{c}\sum_{x,y,z}\sum_{s=2}^{n}\frac{\partial}{\partial x_1}\left[\frac{\partial^2\phi_{s1}}{\partial x_1\partial t_s}-\frac{\partial^2\phi_{s1}}{\partial x_s\partial t_1}\right]\\ &= \frac{1}{c}\sum_{s=2}^{n}\frac{\partial}{\partial t_s}\Box_1^2\phi_{s1}-\frac{1}{c}\frac{\partial}{\partial t_1}\left[\sum_{s=2}^{n}\Box_1\cdot\Box_s\phi_{s1}\right],\end{aligned} \tag{12}$$

where we have written $\Box_1^2 = \nabla_1^2 - \dfrac{1}{c^2}\dfrac{\partial^2}{\partial t_1^2}$,

$$\Box_1\cdot\Box_s = \nabla_1\cdot\nabla_s - \frac{1}{c^2}\frac{\partial^2}{\partial t_1\partial t_s}.$$

Again,

$$\begin{aligned}(\operatorname{curl}\mathbf{H}_1)_x &= \frac{\partial}{\partial y_1}\left[\sum_{s=2}^{n}\left(\frac{\partial^2\phi_{s1}}{\partial x_1\partial y_s}-\frac{\partial^2\phi_{s1}}{\partial y_1\partial x_s}\right)\right]-\frac{\partial}{\partial z_1}\left[\sum_{s=2}^{n}\left(\frac{\partial^2\phi_{s1}}{\partial z_1\partial x_s}-\frac{\partial^2\phi_{s1}}{\partial x_1\partial z_s}\right)\right]\\ &= -\sum_{s=2}^{n}\frac{\partial}{\partial x_s}\Box_1^2\phi_{s1}+\frac{\partial}{\partial x_1}\left[\sum_{s=2}^{n}\Box_1\cdot\Box_s\phi_{s1}\right]+\frac{1}{c}\frac{\partial}{\partial t_1}\left[\sum_{s=2}^{n}\frac{1}{c}\left(\frac{\partial^2\phi_{s1}}{\partial x_1\partial t_s}-\frac{\partial^2\phi_{s1}}{\partial x_s\partial t_1}\right)\right]\end{aligned}$$

or

$$\left(\operatorname{curl}\mathbf{H}_1-\frac{1}{c}\frac{\partial\mathbf{E}_1}{\partial t_1}\right)_x = -\sum_{s=2}^{n}\frac{\partial}{\partial x_s}\Box_1^2\phi_{s1}+\frac{\partial}{\partial x_1}\left[\sum_{s=2}^{n}\Box_1\cdot\Box_s\phi_{s1}\right]. \tag{13}$$

7. The further reduction of identities (12) and (13) thus depends on the structure of the ϕ_{s1}'s. From our gravitational work we are already familiar with the identity

$$\Box_1^2\frac{X_{s1}}{(X_{s1}^2-X_sX_1)^{\frac{1}{2}}} \equiv 0.$$

We now find, as an entirely unforeseen identity, that*

$$\Box_1^2 \frac{1}{(X_{s1}^2 - X_s X_1)^{\frac{1}{2}}} \equiv 0,$$

and accordingly that

$$\Box_1^2 \phi_{s1} \equiv 0. \tag{14}$$

On the other hand, $\Box_1 . \Box_s \phi_{s1} \not\equiv 0$. We find in fact that

$$\Box_1 . \Box_s \phi_{s1} = -\frac{1}{2} \frac{e_s}{Y_1^{\frac{1}{2}} Y_s^{\frac{1}{2}}} \frac{1}{c^2} \frac{2X_{s1} Z_s Z_{s1} - X_s(Z_s Z_1 + Z_{s1} Z_{1s})}{(X_{s1}^2 - X_s X_1)^{\frac{5}{2}}}.$$

That this does not vanish identically is most readily seen by taking the origin at P_s. But $\Box_1 . \Box_s \phi_{s1}$ is a very small scalar, since the leading term in the numerator (the term in $t_s^3 t_1$) vanishes identically. We shall put

$$\sum_{s=2}^{n} \Box_1 . \Box_s \phi_{s1} = a_1, \tag{15}$$

where a_1 is a scalar symmetrical in the suffixes $2, \ldots, n$ and depending moreover on symbols suffixed 1

8. Identities (12) and (13) now reduce to

$$\operatorname{div} \mathbf{E}_1 = -\frac{1}{c} \frac{\partial a_1}{\partial t_1}, \tag{16}$$

$$\left(\operatorname{curl} \mathbf{H}_1 - \frac{1}{c} \frac{\partial \mathbf{E}_1}{\partial t_1} \right)_x = \frac{\partial a_1}{\partial x_1}. \tag{17}$$

Hence
$$\left(\operatorname{curl} \mathbf{H}_1 - \frac{1}{c} \frac{\partial \mathbf{E}_1}{\partial t_1}, \ \operatorname{div} \mathbf{E}_1 \right)$$

constitutes a contravariant 4-vector. Eliminating the scalar a_1 from (16) and the three relations of type (17) by cross-differentiation, and dropping the suffix 1 as no longer necessary, we have the two vector identities

$$\operatorname{grad} \operatorname{div} \mathbf{E} + \frac{\partial}{c\,\partial t} \left[\operatorname{curl} \mathbf{H} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \right] = 0, \tag{18}$$

$$\operatorname{curl} \left[\operatorname{curl} \mathbf{H} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \right] = 0. \tag{19}$$

---

* The number of space-dimensions, 3, plays an essential part in the derivation of this identity.

Rewriting (10) and (11) with the suffix 1 dropped, we have

$$\operatorname{div} \mathbf{H} = 0, \tag{20}$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}. \tag{21}$$

Relations (18), (19), (20), (21) are the field identities replacing Maxwell's equations in the present treatment

9. Maxwell's equations themselves at any point in free space, i.e. at any point where we can put a test-charge e_1 to measure the field, consist of (20) and (21) together with

$$\operatorname{div} \mathbf{E} = 0, \tag{22}$$

$$\operatorname{curl} \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = 0, \tag{23}$$

which correspond in a sense to (18) and (19).

It will be seen that whenever (22) and (23) are satisfied, so are (18) and (19). But if (18) and (19) are satisfied, (22) and (23) are not necessarily satisfied. The right-hand sides of (22) and (23) are in our treatment not in general zero, being given by

$$\operatorname{div} \mathbf{E} = -\frac{1}{c}\frac{\partial a}{\partial t}, \tag{24}$$

$$\operatorname{curl} \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \operatorname{grad} a. \tag{25}$$

Thus our (18) and (19) impose less severe restrictions on E and H than the classical (22) and (23).

We may now recall that Maxwell's two equations (22) and (23) stand on a different footing from the other two (20) and (21), as pointed out by Lorentz (1916). In Maxwell's formulation (20) and (21) represent definite experimental facts, whereas (23) is a pure hypothesis, the hypothesis of the displacement current (in free space), and (22) encounters difficulties, since it would appear to proclaim the non-existence of charges, as (20) does of poles. It is therefore not surprising to find that the two equations of Maxwell which were hypothetical in character, and to which Lorentz drew attention, do not appear in the same form in our deductive treatment. Owing to the smallness of the scalar a, the exact relations (24) and (25) will appear to observation very nearly the same as the usually adopted equations (22) and (23). But, from a strictly logical point of view, (24) and (25), or their consequences (18) and (19), would appear to be preferable.

It must be remembered that in our treatment **H** (and to a lesser extent **E**), which measure the actual field experienced by the moving test-charge, depend on the velocity of the test-charge, whereas in the classical theory **H** and **E** are conceptual quantities existing at every point and independent of the velocity of the test-charge. In our work **H** and **E** nowhere have a meaning until the test-charge is introduced *and its velocity specified*: **H** and **E** are functions of $\mathbf{V}_1$. But identities (18) and (19), like the structurally simpler ones (20) and (21), contain no explicit mention of $\mathbf{V}_1$, and hold good for any value of $\mathbf{V}_1$, provided this is treated as a constant in carrying out partial differentiations.

Another difference is that in the classical theory the properties of the field are deduced from the "field equations" (20), ..., (23), whereas in our treatment (18), ..., (21) are end-products. In the classical theory, to determine the motions of charges an equation of mechanical force (obtained from the Larmor-Lorentz ponderomotive formula) is superposed on the set of field equations; in our treatment we begin with equations of mechanical force, and so *derive* the forms of the super-potentials ϕ_{s1} and ϕ_{1s}. Our treatment seems epistemologically preferable, since to be both logically and epistemologically satisfactory the analysis should begin with the mechanical effects through which alone **E** and **H** can be known. Since no entity exists corresponding to an isolated magnetic pole, it is unsatisfactory to define **H** through the force on a unit pole, as the classical theory does; instead, we have introduced **H** through the force on a moving charge.

Our field identities (18), ..., (21) hold good in t-time and t-measure, and were derived from mechanical equations of motion employing t-measure for the Larmor-Lorentz mechanical force. Our analysis pays in fact due regard to the circumstance that all phenomena, electromagnetic or otherwise, take place in a universe which (in t-measure) is expanding, and which possesses everywhere a local standard of rest; and the analysis respects Mach's principle, according to which all frames of reference employed must be described with regard to the actual distribution of matter in the universe, represented in idealized form by the substratum. The classical treatment, in ignoring the expansion of the universe, can only be supposed to hold good at best locally, and similarly it can only be supposed to hold good for epochs close to the present epoch, since it ignores the distinction between t-measure, for which Lorentz invariance holds good, and τ-measure, which is appropriate to mechanical equations of classical type. There would in fact be grave difficulties in attempting to generalize the classical electromagnetic equations *as they stand* to other epochs than the present one and to distances not small compared with the radius of the universe; for the field equations of the classical theory are stated in t-measure, whilst the Larmor-Lorentz mechanical force formula is correct in τ-measure. Except for $t=\tau=t_0$, the field identities (18), ..., (21) would take a different form in τ-measure. The present method, of beginning systematically in t-measure in flat-space, with the t-form of the mechanical force, appears to be the only way of attacking the problem. It is only in this way that we ascertain *a posteriori* that Maxwell's equations, or rather the modified set to which we have been led, employ t-measure.

10. The substantial agreement of our deductive set (18), ..., (21) with the empirical set (20), . ., (23) is further evidence for the identification of our abstract symbols **E**, **H** with the electric and magnetic intensities of experimental physics; identity (21), for example, guarantees that in spite of our modification of the Biot and Savart law, all effects of electromagnetic induction are properly accounted for. The actual experimental evidence in favour of Maxwell's equations (22), (23) is however indirect, consisting as it does in the fact that the set (20), ..., (23) yield wave propagation of **E** and **H**. We must now enquire whether our set (18), .. , (21) yield wave propagation.

WAVE PROPAGATION

11. Expanding the left-hand side of (19) we have

$$\operatorname{grad}\operatorname{div}\mathbf{H}-\nabla^2\mathbf{H}-\frac{1}{c}\frac{\partial(\operatorname{curl}\mathbf{E})}{\partial t}=0.$$

Substituting from (20) and (21) we get

$$\Box^2\mathbf{H}=0. \tag{26}$$

Thus **H** obeys the wave-equation. Again, (18) may be written

$$\operatorname{curl}\operatorname{curl}\mathbf{E}+\nabla^2\mathbf{E}+\operatorname{curl}\left(\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t}\right)-\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2}=0,$$

whence, using (21), we get $\Box^2\mathbf{E}=0$ (26′)

Thus **E** also obeys the wave-equation. Further, a satisfies the wave-equation. For, returning for the moment to the use of suffixes, we have

$$\Box_1^2 a_1=\Box_1^2\sum_{s=2}^{n}\Box_1.\Box_s\phi_{s1}\equiv\sum_{s=2}^{n}\Box_1.\Box_s(\Box_1^2\phi_{s1})\equiv 0 \tag{26″}$$

in virtue of (14).

RETARDATION EFFECTS

12. The fact that **E**, **H** satisfy wave-equations allows us now to connect an event $(\mathbf{P}_1, t_1)$ with events $(\mathbf{P}_s, t_s)$ at the relevant charges by the usual retardation formulae. All our field identities are explicitly independent of any formulation of this connexion, but in practical calculations of **E** and **H** the connexion is required to link up our abstract scheme with experience. We shall not go into this in detail, but the effect on the fundamental denominator $(X_{1s}^2 - X_1 X_s)^{\frac{1}{2}}$ is worth considering.

The symbols $\mathbf{P}_s$, t_s must refer to the retarded position and epoch of the charge e_s, and accordingly

$$t_1 = t_s + \frac{|\mathbf{P}_1 - \mathbf{P}_s|}{c}.$$

Hence

$$t_1\mathbf{P}_s - t_s\mathbf{P}_1 = t_1(\mathbf{P}_s - \mathbf{P}_1) + \frac{\mathbf{P}_1 \,|\mathbf{P}_1 - \mathbf{P}_s|}{c}.$$

Then since

$$X_{1s}^2 - X_1 X_s = \frac{(t_s\mathbf{P}_1 - t_1\mathbf{P}_s)^2}{c^2} - \frac{(\mathbf{P}_1 \wedge \mathbf{P}_s)^2}{c^4},$$

some vector algebra leads to

$$(X_{s1}^2 - X_s X_1)^{\frac{1}{2}} = \frac{t_1\,|\mathbf{P}_s - \mathbf{P}_1|}{c}\left[1 + \frac{\mathbf{P}_1 . (\mathbf{P}_s - \mathbf{P}_1)}{ct_1\,|\mathbf{P}_s - \mathbf{P}_1|}\right], \tag{27}$$

the radical disappearing. When we choose the origin at the test-particle, we have $\mathbf{P}_1 = 0$ and we get simply

$$(X_{s1}^2 - X_s X_1)^{\frac{1}{2}} = t_1\,|\mathbf{P}_s|/c, \tag{28}$$

where $\mathbf{P}_s$ is the retarded position. Whenever $|\mathbf{P}_1| \ll ct_1$, i.e. for all local phenomena, we have approximately

$$(X_{s1}^2 - X_s X_1)^{\frac{1}{2}} \sim t_1\,|\mathbf{P}_s - \mathbf{P}_1|/c. \tag{29}$$

VECTOR-POTENTIAL

13 Define a covariant 4-vector $(\mathbf{A}_1, A_1)$ by

$$\mathbf{A}_1 = \sum_{s=2}^{n} \frac{\partial \phi_{s1}}{\partial \mathbf{P}_s}, \quad A_1 = \sum_{s=2}^{n} \frac{\partial \phi_{s1}}{c\,\partial t_s}. \tag{30}$$

By (1) and (1′) we then have

$$\mathbf{E}_1 = \frac{\partial A_1}{\partial \mathbf{P}_1} - \frac{\partial \mathbf{A}_1}{c\,\partial t_1} = \operatorname{grad} A_1 - \frac{\partial \mathbf{A}_1}{c\,\partial t_1}, \tag{31}$$

and

$$\mathbf{H}_1 = \operatorname{curl} \mathbf{A}_1. \tag{32}$$

Then
$$\operatorname{div}\mathbf{A}_1-\frac{1}{c}\frac{\partial A_1}{\partial t}=\sum_{s=2}^{n}\Box_1\,.\,\Box_s\phi_{s1}=a_1. \tag{33}$$

Further, by (14),
$$\Box_1^2\mathbf{A}_1=\sum_{s=2}^{n}\frac{\partial}{\partial\mathbf{P}_s}\Box_1^2\phi_{s1}\equiv 0, \tag{34}$$

$$\Box_1^2 A_1=\sum_{s=2}^{n}\frac{\partial}{c\partial t_s}\Box_1^2\phi_{s1}\equiv 0. \tag{34'}$$

We can now as before suppress the suffix 1, and employ a covariant vector-potential $(\mathbf{A}, A_t)$ giving the covariant 6-vector $(\mathbf{H}, \mathbf{E})$ by the formulae

$$\mathbf{E}=\operatorname{grad}A_t-\frac{\partial\mathbf{A}}{c\partial t},\quad \mathbf{H}=\operatorname{curl}\mathbf{A}, \qquad (35),\ (35')$$

$$\operatorname{div}\mathbf{A}-\frac{\partial A_t}{c\,\partial t}=a, \tag{36}$$

$$\Box^2\mathbf{A}=0,\quad \Box^2 A_t=0. \tag{37}$$

These formulae bring out the difference between our treatment and the classical formulation. In the classical theory there is initially an arbitrariness in the vector-potential, an arbitrariness which is conveniently removed by imposing the condition*

$$\operatorname{div}\mathbf{A}-\frac{\partial A_t}{c\,\partial t}=0.$$

In our treatment there is no such arbitrariness, and the right-hand side of the last equation is not zero, but, by (36), just a, which is a definite function of the charges present. It will be seen that whilst the formal possibility of representing **E** and **H** by a vector-potential arises from our original definitions of **E**, **H** in terms of the super-potentials ϕ_{s1}, the wave propagation property of **A** depends on the detailed structure of the ϕ_{s1}'s. We nowhere consider "fields" in the abstract, but only such fields as can arise from definite distributions of charge. The vector-potential is an intermediary between the super-potentials ϕ_{s1} and the scalar a.

It is clear that A_t or A_1 is the negative electrostatic potential for a "stationary" field. It should therefore stand in a simple relation to the interaction energy $\sum_{s=2}^{n}\Phi_{s,1}$ of e_1 with the other charges in the field. For a set of stationary charges we find by actually carrying out the differentiations and choosing the origin at P_1 that

---

* See, e.g. Lorentz (1916, p. 239). Our sign convention for A_t differs from his, but ours brings out the analogy between the space- and time-co-ordinates in the expressions for **E** and **H**.

$$\sum_{s=2}^{n} \Phi_{s,1} = \frac{t_1}{t_0}(-e_1 A_1) = \sum_{s=2}^{n} \frac{e_1 e_s}{|\mathbf{P}_s|}\frac{t_s}{t_0}. \tag{37'}$$

The existence of this relation is a general check on the correctness of our formulae for ϕ_{s1} and $\Phi_{s,1}$.

PROPAGATION OF LIGHT

14. The fact that $\mathbf{E}, \mathbf{H}, (\mathbf{A}, A_t)$ and a satisfy the wave-equation in t-measure now shows that our treatment is compatible with the usual electromagnetic theory of light, provided t-measure is used. This suggests that a photon is propagated with a constant frequency when the frequency is reckoned on the t-scale. This is in accordance with an investigation due to Whitrow (1936), who discussed the photon by purely kinematic methods without appeal to electromagnetic theory, and established the proportionality of photon-energy to photon-frequency on the t-scale. We shall show in an Appendix that a Bohr atom has a constant energy W on the t-scale, and hence that it emits and absorbs a frequency which is constant on the t-scale. We have to reconcile this with the τ-measure employed in ordinary physics. When t-measure is employed, the red-shift in the spectrum of a distant, receding fundamental particle appears as a Doppler shift. We shall now examine how the same phenomenon is regarded in τ-measure.

Let n be the measure of a photon-frequency on the t-scale, ν the measure of the frequency of the same photon *at the same epoch* on the τ-scale. Then since measures Δt, $\Delta\tau$ of the same small interval of time on the two scales are connected by

$$\frac{\Delta t}{t} = \frac{\Delta\tau}{t_0},$$

we have

$$\nu = \frac{1}{\Delta\tau} = \frac{t}{t_0}\frac{1}{\Delta t} = \frac{t}{t_0}n. \tag{38}$$

Now let a photon of frequency n_0 on the t-scale be emitted by a distant fundamental particle (or extra-galactic nebula) at epoch $t_0 = \tau_0$, and be received by the observer at epoch t_1 or τ_1, where

$$\tau_1 = t_0 \log\frac{t_1}{t_0} + t_0. \tag{39}$$

Wave-lengths l or λ on the two scales are connected with the corresponding frequencies by $\lambda\nu = c = ln$. By (38) the emitted frequency satisfies $\nu_0 = n_0$, and $\lambda_0 = c/\nu_0 = c/n_0$. On the τ-scale, the fundamental particles appear as relatively stationary, and there is no Doppler shift. Since frequencies are transmitted unchanged, the receiving observer, who has made no allowance for Doppler shift, will reckon the received wave-length as λ_0.

Now consider the similar atom at the receiver. This will be absorbing the same frequency on the t-scale, and so the observer will call this frequency n_1 or ν_1 on the two scales, where $n_1 = n_0$, and by (38)

$$\nu_1 = \frac{t_1}{t_0} n_1. \tag{40}$$

Hence he will reckon the wave-length shift-ratio as

$$s = \frac{\text{wave-length calculated as received}}{\text{wave-length absorbed by a similar atom at epoch of reception}}$$

$$= \frac{\lambda_0}{\lambda_1} = \frac{\nu_1}{\nu_0} = \frac{t_1 n_1}{t_0 n_0} = \frac{t_1}{t_0}, \tag{41}$$

or by (39)

$$s = e^{(\tau_1 - t_0)/t_0}. \tag{42}$$

But if Λ is the (fixed) distance of the emitter on the τ-scale,

$$\tau_1 = \tau_0 + \Lambda/c = t_0 + \Lambda/c.$$

Hence

$$s = e^{\Lambda/ct_0}. \tag{43}$$

This is precisely the value found by a t-calculation directly from the standard Doppler-effect formula in an earlier paper (Milne 1937, equation (22)). We see that on the t-scale the observer calculates the red-shift as a pure Doppler effect; on the τ-scale, by (40), he considers the frequency of the standard atom at himself as steadily increasing, so that light emitted by a distant atom at an earlier epoch has apparently a lower frequency and so a longer wave-length at the moment of reception.

We know from earlier work that if $|\mathbf{P}_1|$ is the measure of a distance on the t-scale at epoch t_1, $|\mathbf{\Pi}_1|$ the measure of the same distance on the τ-scale at the same epoch, then

$$|\mathbf{P}_1| = \frac{t_1}{t_0} |\mathbf{\Pi}_1|. \tag{44}$$

If $|\mathbf{P}_1|$ is constant, $|\mathbf{\Pi}_1|$ decreases with epoch inversely as t_1. Thus the foregoing shows that a wave-length λ_1 on the τ-scale behaves just like a "distance".

If the standard metre-rod is defined as a fixed number of wave-lengths, it should accordingly be decreasing on the τ-scale, constant on the t-scale. But if it is defined as a material rod, consisting of a fixed number of atoms held together by electrostatic forces, we shall see in the Appendix that it should be increasing on the t-scale, constant on the τ-scale. It follows that the number of wave-lengths contained in the material standard metre at any instant should be increasing at the rate of about one part in 2×10^9 per year.

COSMOLOGICAL CONSIDERATIONS

15. It is relevant to the above calculation to point out here that the concurrent use of the t- and τ-scales in this manner completely removes the difficulties encountered by Hubble in his recent book (1937). Hubble considered that either the red-shift must be attributed to a motion of recession, or else, if the nebulae are considered as stationary, it must be attributed to some entirely unknown cause new to physics. We now see that these are not two *distinct* possible phenomena, but different descriptions of the *same* phenomenon according to the scale of time adopted. In further confirmation, it may be pointed out that Hubble's observations disclosed a density-distribution of nebulae increasing outwards if recession is adopted, and a homogeneous distribution if recession is denied. This is just what is predicted on the present treatment. In t-measure, corresponding to recession, the substratum gives an observable "world-picture" with a density-distribution increasing outwards; in τ-measure it was shown (Milne 1937, p. 181) that in the hyperbolic space appropriate to this measure the density-distribution is strictly homogeneous.

ENERGY AND ELECTROMAGNETIC RADIATION

16. In calculating the energy of a system of charges two points must be attended to. First, there is now no such thing as a test-charge, and e_1 is an intrinsic part of the complete system, $e_1, \ldots, e_n$. Secondly, to introduce a measure of "the energy of a system at an instant", we must adopt some standard of simultaneity for the different charged particles in the experience of the given observer. Having fixed the observer O, we fix as his standard of simultaneity $t_1 = t_2 = \ldots = t_n = t$. This is not an invariant relation, as the set of events selected varies from observer to observer.* Other observers will

---

* The same difficulty does not appear in τ-measure, since τ is an invariant, behaving as a Newtonian time.

choose "simultaneous" events differently, although using the same convention. Thus although expression (8) for the energy Φ is invariant for all observers for a given set of events $(\mathbf{P}_1, t_1) \dots (\mathbf{P}_n, t_n)$, it takes different numerical values for different observers on inserting $t_1 = t_2 = \dots = t_n = t$, since the particles will be in different places P_s for the same value of t for different observers. This accounts for the fact that though (6) is derived from invariant relations, and holds good identically even without making use of "retardation" formulae, it is not itself an invariant relation. The right-hand side represents the negative rate of radiation of the system, but until "energy at an instant" has been defined, the phrase "rate of radiation" is meaningless.

It is to be noticed that we have nowhere assumed the conservation of energy. Energy can only be defined in terms of mechanical work; the terms $m_s c^2 \xi_s^{\frac{1}{2}}$ were shown in earlier papers to represent the accumulation of mechanical work performed on the particles of masses m_s, positions P_s, epochs t_s and velocities $\mathbf{V}_s$ by the total external forces acting together with the pull of the substratum itself, the term Φ was then added to make the total stationary when the accelerations $\dot{\mathbf{V}}_s$ vanish. But the right-hand side of (6) may *a priori* be negative or positive. It would be irrelevant to ask "Where does the energy come from?" when the right-hand side is positive; energy is a construct, not a concept with which we have direct acquaintance.

17. We now notice that our abandonment of the concept of an electromagnetic field as existing at a given epoch at a given point independently of the velocity of the test-charge used to measure it, and its replacement by an analysis making $\mathbf{E}$ and $\mathbf{H}$ functions of the velocity $\mathbf{V}$ of the test-charge, render it impossible to represent an electromagnetic field as a state of the "aether", even for a fixed observer. And since $\mathbf{E}$ and $\mathbf{H}$ do not exist as measurable quantities until the velocity $\mathbf{V}$ of the test-charge is stated, $(\mathbf{E}^2 + \mathbf{H}^2)/8\pi$ integrated through space has by itself no physical meaning. Hence in our treatment we cannot expect electromagnetic energy to be measurable by a volume integral. Instead, we measure it by (8). This immediately avoids the old difficulty of the infinity of the self-energy of a point-charge on the classical theory, since in (8) energy is attributed only to *pairs* of charges.

As a consequence we have no Poynting theorem; (6) replaces it. In the classical derivation of Poynting's theorem, the rate of performance of work on the charges present is calculated and converted into volume and surface integrals by replacing the charge-density ρ by $(\operatorname{div} \mathbf{E})/4\pi$. In our treatment there is no ρ, but only point-charges, and we have no relation $\operatorname{div} \mathbf{E} = 4\pi\rho$

at our disposal (cf. (22)). Consequently the mathematical manipulations leading to Poynting's theorem are not possible.

But (6) fulfils all the duties of Poynting's theorem. It would *prima facie* resemble it more closely if we made the hypothesis of the purely electromagnetic origin of mass, which would be equivalent to putting each m_s in (6) equal to zero. But this, besides being an *ad hoc* hypothesis, is not desirable on logical grounds, for inertial mass emerged in our kinematic treatment at a much earlier stage than electromagnetic phenomena—which indeed were only capable of discussion after gravitation had been dealt with. Since gravitational mass was proved equal to inertial mass m_s, to put the inertial masses m_s equal to zero would exclude the possibility of attributing gravitational effects to whatever emerged in the subsequent treatment as electromagnetic mass.

Classical electromagnetism results ultimately in a conflict with experience: it predicts that a set of charges moving with accelerations $\dot{\mathbf{V}}_s$ should radiate energy at a rate proportional to $\left(\sum_s e_s \dot{\mathbf{V}}_s\right)^2$.* An atom in a stationary state contains accelerated charges but is not radiating, in contradiction, as is well recognized, with the prediction of the classical theory. Now the classical result is obtained by applying Poynting's theorem, which essentially involves a belief in continuous distributions of charge of finite space-density; in our treatment, which gives a rigorous account of point-charges, instead of Poynting's theorem we have (6), which makes the rate of change of energy a *linear*, not a *quadratic*, function of the accelerations.

It would make the present paper too long to attempt an exhaustive discussion of the right-hand side of (6) and its comparison with the classical theory. I shall however consider briefly a particular case

Take the case where only two charged particles are present. Then we have

$$\Phi = \Phi_{12} = \frac{1}{2}\frac{e_1 e_2}{ct_0}\,\frac{Z_{12}Z_{21} + Z_1 Z_2}{(X_{12}^2 - X_1 X_2)^{\frac{1}{2}} Y_1^{\frac{1}{2}} Y_2^{\frac{1}{2}}}.$$

Hence

$$\frac{1}{\Phi}\frac{\partial \Phi}{\partial \mathbf{V}_1} = \frac{\mathbf{V}_1/c^2}{Y_1} - \frac{(Z_{12}\mathbf{P}_2 + Z_2\mathbf{P}_1)/c^2}{Z_{12}Z_{21} + Z_1 Z_2}.$$

Hence

$$\frac{\partial \Phi}{\partial \mathbf{V}_1} \sim \frac{\Phi}{c}\left[\frac{\mathbf{V}_1/c}{Y_1} - \frac{t_1\mathbf{P}_2 + t_2\mathbf{P}_1}{2ct_1 t_2}\right].$$

Putting $t_1 = t_2 = t$,

$$\frac{\partial \Phi}{\partial V_1} \sim \frac{\Phi}{c}\left[\frac{\mathbf{V}_1/c}{Y_1} - \frac{\mathbf{P}_1 + \mathbf{P}_2}{2ct}\right].$$

---

* Larmor 1900, Abraham 1914, Richardson 1916.

The second term in the square bracket is negligible in ordinary experience. Hence approximately

$$\dot{\mathbf{V}}_1 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_1} + \dot{\mathbf{V}}_2 \cdot \frac{\partial \Phi}{\partial \mathbf{V}_2} \sim \frac{\Phi}{c^2}\left[\frac{\mathbf{V}_1 \cdot \dot{\mathbf{V}}_1}{Y_1} + \frac{\mathbf{V}_2 \cdot \dot{\mathbf{V}}_2}{Y_2}\right]. \tag{45}$$

By (6), this should be the negative rate of radiation of energy.

Now let one of the charges, e_2, be the nucleus of an atom at rest, so that $\mathbf{V}_2 = 0$, and let the other, e_1, be an electron moving in a "circular" orbit* round e_2. Then the acceleration $\dot{\mathbf{V}}_1$ is perpendicular to the velocity $\mathbf{V}_1$, and $\mathbf{V}_1 . \dot{\mathbf{V}}_1 = 0$. Accordingly by (6) and (45) the sum of the mechanical and electromagnetic energies remains constant, and the system is in a non-radiating, i e. stationary, state. This is in accordance with the hypotheses of the quantum theory, in contrast to classical electromagnetic theory. It will be noticed that we have calculated the energy exactly as in Bohr's original calculation, treating the electron as a massive point-charge.

For an elliptic orbit, $\Phi \propto 1/|\mathbf{r}|$, $\mathbf{V}_1 = \dot{\mathbf{r}}$, $\dot{\mathbf{V}}_1 = \ddot{\mathbf{r}} \propto -\mathbf{r}/|\mathbf{r}|^3$, and the negative rate of radiation for $|\dot{\mathbf{r}}| \ll c$, is proportional to

$$-\frac{1}{|\mathbf{r}|}\frac{\dot{\mathbf{r}} . \mathbf{r}}{|\mathbf{r}|^3}$$

$$= \frac{1}{2}\frac{d}{dt}\frac{1}{\mathbf{r}^2}.$$

The mean value of this in an orbit is zero, and so an elliptic orbit is also a stationary state. The actual "energy" fluctuates.

The difficulty encountered on the classical theory, of how radiationless periodic systems containing accelerated charges can exist, is removed on the present theory as far as these examples go. Actual radiation must therefore be associated with interruptions of strictly periodic motion. An obvious example in which our treatment gives non-zero radiation is the case of a cathode ray particle entering matter, when the high speed electron has a linear deceleration. In this case $\dot{\mathbf{V}}_1$ is in the opposite direction to $\mathbf{V}_1$, and the term $\mathbf{V}_1 . \dot{\mathbf{V}}_1$ makes a non-zero contribution. Hence such a particle should radiate, as it does in the form of X-rays.

Since the classical theory appears to give the right order of magnitude for the radiation when radiation is occurring, it is desirable to compare the actual rates of radiation on the two theories. To illustrate orders of magnitude, consider the following particular case. If an electron $-e$ is in motion about a nucleus $+e$ with acceleration f, velocity v, at distance r, the classical rate of radiation is $b_1 = \frac{2}{3}e^2f^2/c^3$. On our treatment the rate of radiation depends on the scalar product of the vectors corresponding to f and v, and fluctuates with zero as its mean value over the orbit; its numerical value b_2 at any instant, by (45), lies between 0 and $vf|\Phi|/c^2$. Since $|\Phi| = e^2/r$, the ratio b_2/b_1 lies numerically between zero and something less than cv/rf. Now f is comparable with v^2/r. Hence b_2/b_1 lies numerically between zero and something less than, but of the order of, c/v. For a nearly circular orbit it remains small, but it increases with the ellipticity. For a highly elliptical orbit, the classical radiation occurs principally near perihelion, and here v approaches the order of magnitude of c. This is sufficient to show that whilst the radiation is differently distributed over the orbit in the two cases, the orders of magnitude of the absolute values are not widely different. On the present treatment the total radiation over a complete orbit is zero, so that possibly actual radiation is to be associated either with the incompleted portions of orbits interrupted by a transition or with the transition itself. More detailed calculations would be required to investigate this, but space forbids. It is at least satisfactory that completed orbits give zero net radiation. If discrepancies are found in the actual amount of the radiation, it should be remembered that we are treating electrons as points, whilst the classical calculation, which uses Wiechert's formula for the retarded potentials, depends essentially on treating an electron as a small region of finite charge-density (Richardson 1916, p. 244), though it is incapable of dealing with the forces (necessary to hold the electron together) which are implied by this view. Whatever view we take of the structure of ultimate particles, either they are for purposes of analysis infinitely divisible or else they reduce to a system of singularities. The present treatment should give the theory of pure singularities, whether the electron is an ultimate singularity or not.

---

* The secular change of radius with t in t-measure is of course of no significance in the present calculation, which is concerned with short time-intervals.

18. In the present treatment "energy" is not conserved from moment to moment, and the field itself is not the seat of the energy, but a mechanism for transferring energy from one particle to another. The theory thus amounts to an embodiment in analysis of the views put forward in 1924 by Bohr, Kramers and Slater, though in origin it is quite different, being a logical construct based on the hypothesis of zero radiation for non-accelerated systems and on a proper epistemological treatment of **E** and **H** and their mechanical effects.

GRAVITATION AND ELECTROMAGNETISM

19. I conclude the paper by showing how to handle in relativistic form problems in which gravitational and electromagnetic fields occur together. The only technical difficulty is the construction of an appropriate measure of external force. We use the t-scale of time.

Let χ be the total gravitational potential due to condensations in the substratum, as constructed in previous papers. (The field of the substratum itself is of course taken into account through the form of the equations of motion.) Let $\mathbf{E}$, $\mathbf{H}$ be the electric and magnetic intensities at a particle of mass m and charge e when its position vector is $\mathbf{P}$, epoch t and velocity $\mathbf{V}$. Then we seek to determine a scalar α such that the external force $(\mathbf{F}, F_t)$ can be represented by

$$\mathbf{F} = -\frac{\partial\chi}{\partial\mathbf{P}} + \frac{Z}{Y^{\frac{1}{2}}}\frac{e}{t_0}\left[\mathbf{E} + \frac{\mathbf{V}\wedge\mathbf{H}}{c}\right]\frac{1}{Y^{\frac{1}{2}}} + \alpha\frac{\mathbf{V}}{Y^{\frac{1}{2}}}, \tag{46}$$

$$F_t = \frac{\partial\chi}{c\,\partial t} + \frac{Z}{Y^{\frac{1}{2}}}\frac{e}{t_0}\left[\frac{\mathbf{E}.\mathbf{V}}{c}\right]\frac{1}{Y^{\frac{1}{2}}} + \alpha\frac{c}{Y^{\frac{1}{2}}}. \tag{46'}$$

When these are inserted in the equations of motion (3), (3′) of Part I, the energy equation (7) (Part I) yields

$$\frac{1}{Y^{\frac{1}{2}}}\frac{d\chi}{dt} + \alpha c^2 = \frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(mc^2\xi^{\frac{1}{2}}). \tag{47}$$

The other energy equation (6) (Part I), when we use the fact that χ is homogeneous in $\mathbf{P}$ and t of degree zero, yields

$$\frac{e}{t_0}\left[\frac{\mathbf{E}.(\mathbf{V}t-\mathbf{P})}{Y^{\frac{1}{2}}} - \frac{\mathbf{P}\wedge\mathbf{V}.\mathbf{H}}{cY^{\frac{1}{2}}}\right] = \frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}[mc^2\xi^{\frac{1}{2}} + \chi]. \tag{48}$$

When $\mathbf{E}=0$, $\mathbf{H}=0$ we get from (47) and (48)

$$\alpha = \frac{2}{Y^{\frac{1}{2}}}\frac{d}{dt}(m\xi^{\frac{1}{2}}),$$

the standard value for α for a purely gravitational field. When $\chi=0$, we get

$$\alpha = \frac{1}{Y^{\frac{1}{2}}}\frac{d}{dt}(m\xi^{\frac{1}{2}}),$$

as in (9), Part I. The contrast between the last two relations brings out the difference between gravitational and electromagnetic fields.

Equation (48) may be written in the form

$$\frac{d}{dt}[mc^2\xi^{\frac{1}{2}}+\chi] = e\frac{t}{t_0}\left[\mathbf{E}+\frac{\mathbf{V}\wedge\mathbf{H}}{c}\right].\left(\mathbf{V}-\frac{\mathbf{P}}{t}\right),$$

and so exhibits the rate of change of kinetic energy plus gravitational energy as the rate of performance of work by the Larmor-Lorentz force in pushing the particle with velocity $(\mathbf{V}-\mathbf{P}/t)$ relative to the local standard of rest—the whole modified by the secular factor (t/t_0).

20. When the gravitational and electromagnetic fields are due to a given set of massive charged particles, we use a separate epoch co-ordinate t_s for each particle m_s, and a separate coefficient α_s. The gravitational potential is $\chi = \sum\limits_{r,s}\chi_{r,s}$. Forming the energy equations for each particle in the presence of the remainder, we find the analogue of (47) to be

$$\frac{1}{Y_s^{\frac{1}{2}}}\left[\frac{\partial\chi}{\partial t_s}+\mathbf{V}_s.\frac{\partial\chi}{\partial\mathbf{P}_s}\right]+\alpha_s c^2 = \frac{1}{Y_s^{\frac{1}{2}}}\frac{d}{dt_s}(m_s c^2\xi_s^{\frac{1}{2}}), \tag{49}$$

and the analogue of (48) to be

$$\frac{e_s}{t_0}\left[\frac{\mathbf{E}_s.(\mathbf{V}_s t_s-\mathbf{P}_s)}{Y_s^{\frac{1}{2}}}-\frac{\mathbf{P}_s\wedge\mathbf{V}_s.\mathbf{H}_s}{cY_s^{\frac{1}{2}}}\right] = \frac{1}{Y_s^{\frac{1}{2}}}\left[\frac{\partial\chi}{\partial t_s}+\mathbf{V}_s.\frac{\partial\chi}{\partial\mathbf{P}_s}\right]+\frac{1}{Y_s^{\frac{1}{2}}}\frac{d}{dt_s}(m_s c^2\xi_s^{\frac{1}{2}}). \tag{50}$$

Multiplying the last equation by $Y_s^{\frac{1}{2}}dt_s/dt$ and adding n similar equations we get

$$\frac{d}{dt}\left[\sum_{s=1}^{n} m_s c^2\xi_s^{\frac{1}{2}}\right]+\frac{d\chi}{dt} = \Sigma[\text{Rate of performance of electromagnetic work}].$$

The right-hand side may be reduced as in Part I, and we get eventually as the grand energy equation

$$\frac{d}{dt}\left[\sum_{s=1}^{n} m_s c^2\xi_s^{\frac{1}{2}}+\chi+\Phi\right] = \sum_{s=1}^{n}\dot{\mathbf{V}}_s.\frac{\partial\Phi}{\partial\mathbf{V}_s}. \tag{51}$$

The right-hand side is the negative rate of radiation, which is obtained at the joint expense of the mechanical, gravitational and electromagnetic energies. In the last equation it is supposed that we have taken

$$t_1 = t_2 = \ldots = t_n = t.$$

The equations preceding this are invariant in form on transformation from any fundamental observer to any other, and the whole set constitutes a unified relativistic treatment of gravitation and electromagnetism.

REFERENCES

Abraham 1914 "Theorie der Elektrizität", 2, 63. Leipzig.
Bohr, Kramers and Slater 1924 *Phil. Mag.* 19, 436.
Hubble 1937 "Observational Approach to Cosmology." Oxford.
Larmor 1900 "Aether and Matter", p. 227. Cambridge.
Lorentz 1916 "Theory of Electrons", pp. 12, 239. Leipzig.
Milne 1937 *Proc. Roy. Soc.* A, 159, 171.
Richardson 1916 "Electron Theory of Matter", p. 258. Cambridge.
Whitrow 1936 *Quart. J. Math* 7, 271.

APPENDIX

The Bohr atom. Dirac's cosmological relation

1. As an example of the foregoing results, a simple Bohr atom will be treated according to the t-dynamics and t-electrodynamics.

Consider an electron of charge $-e$ moving in the neighbourhood of a massive nucleus of charge $+Ze$. The electrostatic force of attraction in t-measure, by (46) and (43) (Part I), is $-(t/t_0)\,Ze^2/r^2$. The vector equation of motion is equivalent to an equation of radial acceleration and an integral of angular momentum. The former is

$$m(\ddot{r}-r\dot{\theta}^2) = -m\frac{r-\dot{r}t}{t^2}-\frac{Ze^2}{r^2}\frac{t}{t_0}, \tag{52}$$

neglecting relativistic refinements and assuming the nucleus at local rest. The system is "semi-isolated" (Milne 1937*a*, p. 326), and its angular momentum is accordingly proportional to t. We therefore equate its angular momentum to $(zh_0/2\pi)(t/t_0)$, where z is the azimuthal quantum number and h_0 is the value of Planck's constant corresponding to the choice t_0 of our normalization constant, thus $zh_0/2\pi$ is the angular momentum at epoch $t=t_0$. In doing this we are of course appealing to quantum rules outside our strictly kinematic synthesis. We then have

$$mr^2\dot{\theta} = \frac{zh_0}{2\pi}\frac{t}{t_0}. \tag{53}$$

As shown in an earlier paper (Milne 1937*c*, p 17) in connexion with spiral nebulae, the solution of (52) and (53) corresponding to circular motion in the τ-dynamics is

$$r = r_0\frac{t}{t_0}, \quad \dot{\theta} = \frac{zh_0}{2\pi m r_0^2}\frac{t_0}{t}, \tag{54}$$

where

$$r_0 = \frac{z^2h_0^2}{4\pi^2 mZe^2}. \tag{55}$$

In t-measure the orbits are equiangular spirals, and $r \propto t$. This is consistent with the description of the substratum, for the substratum had zero linear dimensions at $t=0$, and necessarily always contains the atom. On the τ-scale the orbits are circular, of constant radius ρ given, by (35′), Part I, by

$$\rho = r\frac{t_0}{t} = r_0. \tag{56}$$

The atom thus behaves as a "rigid" body on the τ-scale, and a material rod built up of such atoms held together by electrostatic mechanical forces will also be "rigid" on the τ-scale. It will therefore be expanding uniformly on the t-scale. It follows that if the nucleus of a distant "receding" nebula were imagined joined to ourselves by a material "rigid" rod, the end of the rod would maintain coincidence with the nebula.* This is in agreement with the "stationariness" of the nebulae as reckoned on the τ-scale (Milne 1937*b*, p. 178). Moreover, a material "rigid" pendulum will keep τ-time.

In spite of the variation of t-radius, the energy of the atom remains constant. In t-measure, the kinetic energy T is given by

$$T = \tfrac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) = \tfrac{1}{2}mr_0^2\left[\frac{1}{t_0^2} + \left(\frac{t}{t_0}\right)^2\dot{\theta}^2\right].$$

The first term in the square bracket, arising from the small outward component of the spiral motion, is relatively negligible, and so approximately

$$T = \frac{2\pi^2 m Z^2 e^4}{z^2 h_0^2}.$$

By formula (33), Part I, the electrostatic energy is

$$\Phi = -\frac{Ze^2}{r}\frac{t}{t_0} = -\frac{4\pi^2 m Z^2 e^4}{z^2 h_0^2}.$$

Hence the total energy is given by

$$W = T + \Phi = -\frac{2\pi^2 m Z^2 e^4}{z^2 h_0^2} = \text{const.}, \tag{57}$$

as on the Bohr theory.

When the atom undergoes a transition, the energy lost is transferred by the mechanism of the field-changes to some external system, and can be reckoned in the meantime as the energy of the liberated photon. If ΔW is its energy, it was shown by Whitrow (1936) by kinematic arguments that its frequency n on the t-scale satisfies

---

* This question was originally suggested to me by Dr F. Simon.

$$\frac{\Delta W}{n} = \text{universal constant.} \tag{58}$$

Assuming this constant to be h_0, we get the Bohr formula, with the frequency reckoned on the t-scale.

2. A referee has drawn my attention to the circumstance that the above calculations appear at first sight to violate the correspondence principle. For by (54), the rotational frequency $\dot{\theta}/2\pi$ in t-measure is proportional to $1/t$, whilst the frequency n of the photon in t-measure is constant. The correspondence principle asserts that for z large and $\Delta z = 1$, the two frequencies should be equal. The explanation arises from the fact that the correspondence principle is an empirical relation between the *dynamical* frequencies of a mechanical system and the *optical* frequencies Rational dynamics as used in physics is stated in terms of the τ-scale of time, and the calculated mechanical frequencies are expressed on the τ-scale. On the other hand, we have shown in Part II that the electromagnetic elements satisfy the wave-equation on the t-scale, and the frequency of an electromagnetic vibration will be propagated as a constant on the t-scale; moreover the purely kinematic considerations advanced by Whitrow show that the frequency of a photon is constant on the t-scale. The two scales and all corresponding measures agree when the normalization constant t_0 is chosen to be the present value of t, but for other values of t we shall expect the correspondence principle to take the form that in specified circumstances *the mechanical frequency on the τ-scale is equal to the optical frequency on the t-scale.*

In this form, the correspondence principle is satisfied by the simple Bohr atom just considered For if $w\,(= \dot{\theta}/2\pi)$ is the mechanical frequency on the t-scale, ω the same frequency on the τ-scale, then

$$\omega = w\frac{t}{t_0} = \frac{\dot{\theta}}{2\pi}\frac{t}{t_0} = \frac{4\pi^2 m Z^2 e^4}{z^3 h_0^3} = \text{const.,} \tag{59}$$

by (54) and (55) Hence for z large, $\Delta z = 1$,

$$\omega \sim \frac{\Delta W}{h_0} \sim n. \tag{60}$$

This puts the correspondence principle in a very remarkable light In connecting mechanical with optical frequencies it is the one bridge connecting the two scales of time. With the interpretation we have given, it puts in evidence a fundamental distinction between matter and radiation. Natural mechanical time-keepers keep τ-time; natural optical time-keepers keep t-time. On the t-scale, radiation in the universe is to be considered as reddened by recession; on the τ-scale, atoms in a stationary universe emit and absorb radiation of a frequency $\nu = n(t/t_0)$ which becomes faster or bluer as the atom ages. It is not a fanciful speculation to see in the interplay of radiation keeping t-time with matter obeying the classical laws of mechanics on the τ-scale a phenomenon giving rise to the possibility of change in the universe *in time*, and so an origin for the action of evolution in both the inorganic and organic universes

3. We now consider Dirac's cosmological relation (1937) in the light of the treatment of electrodynamics given in Parts I and II. The measure e of a charge was introduced *via* a normalization constant t_0, such that on the t-scale the number e agrees with the ordinary electrostatic measure at the epoch $t = t_0$. The number e is of course a constant, but we should write it as e_0 to indicate its dependence on t_0. If we change the value of the normalization constant, then the number e_0 is changed. To calculate this, we recall that the mechanical force between two equal charges e_0 on the t-scale, when at distance r at time t, is

$$\frac{t}{t_0}\frac{e_0^2}{r^2}$$

by (47), Part I This must have the same value whatever value is chosen for t_0, and so if t_0 is changed to t_1, e_0 is changed to e_1, where

$$\frac{e_0^2}{t_0} = \frac{e_1^2}{t_1} \tag{61}$$

Thus the ratio e_0^2/t_0 is independent of the value used for t_0

Similar considerations apply to Planck's constant. We have defined h_0 as 2π times the angular momentum in the first Bohr orbit at $t = t_0$. At time t the angular momentum is then $(h_0/2\pi)(t/t_0)$, and this must be independent of the choice of normalization constant Hence if h_1 corresponds to t_1,

$$\frac{h_0}{t_0} = \frac{h_1}{t_1}. \tag{62}$$

It must not be supposed that h or e change with the time.* They are constants in any given calculation, for all t, but their values depend on t_0.

---

* The possibility of h varying with the time has been considered by Chalmers and Chalmers (1934), Nernst (1935) and others, but the present treatment is quite different.

It now follows that the fine-structure constant* $hc/2\pi e^2$ is independent of the normalization constant t_0, as it should be, since it is a pure number. We have in fact by (61) and (62)

$$\frac{h_0}{e_0^2} = \frac{h_1}{e_1^2}. \tag{63}$$

Similarly the energy W of the Bohr atom just considered is independent of choice of t_0, since it involves e_0^4/h_0^2. On the other hand, the τ-measure of the radius, namely r_0, involves h_0^2/e_0^2, and so is proportional to the value chosen for t_0. This must be carefully distinguished from the circumstance that the t-measure of the radius increases proportionally to t. The same situation arises with regard to the τ-measure of the length of a "rigid" rod For $r = \rho(t/t_0)$, and so $\rho \propto t_0$. The number of wave-lengths contained in a given material rod is independent of t_0; for the wave-length l, from $ln = c$, $n = \Delta W/h_0$, is also proportional to t_0.

We are now in a position to consider Dirac's relation. The gravitational attraction between a proton of mass m_p and an electron of mass m_e gives rise to a mechanical force F_g given in t-measure by

$$F_g = \frac{\gamma m_p m_e}{r^2}, \tag{64}$$

where

$$\gamma = c^3 t/M_0, \tag{65}$$

and M_0 is the mass of the (fictitious) homogeneous universe. Its value determined from $\gamma_0 = 6{\cdot}66 \times 10^{-8}$ at the present epoch $t = t_0 = 2 \times 10^9$ years $= 0{\cdot}6 \times 10^{17}$ seconds is $M_0 = 2{\cdot}4 \times 10^{55}$ grams. (The same value is given by $M_0 = \frac{4}{3}\pi(ct_0)^3\rho_0$, where $\rho_0 = 10^{-27}$ gram. cm.$^{-3}$ is the present mean density of the matter in the universe near ourselves.) The ratio α of M_0 to m_p has the value

$$\alpha = \frac{M_0}{m_p} = 1{\cdot}5 \times 10^{79}. \tag{66}$$

The electrostatic attraction between the proton and electron gives rise to a mechanical force F_e given in t-measure by

$$F_e = \frac{t}{t_0}\frac{e_0^2}{r^2}, \tag{67}$$

where e_0 is now the charge on the electron. Hence the ratio β of the electrostatic to the gravitational attraction is given by

$$\beta = \frac{F_e}{F_g} = \frac{e_0^2 M_0}{c^3 t_0 m_p m_e} = 2{\cdot}3 \times 10^{39}. \tag{68}$$

---

* Professor Dirac kindly directed my attention to this point.

Dirac drew attention to the fact that α is approximately the square of β. Assuming the relation to be exact we get

$$\frac{M_0}{m_p} = \left[\frac{c^3 t_0 m_e}{e_0^2}\right]^2. \tag{69}$$

From this Dirac concluded that $M_0 \propto t^2$, and inferred a continual creation of matter in the universe. But such an inference is not justified. Relation (69), if true, gives M_0 as a certain multiple of $(t_0/e_0^2)^2$, which is independent of the choice of normalization constant t_0. The value of t_0 for the t-unit of charge to agree with the present electrostatic unit is the present value of t, but t_0 is a *constant* in all calculations. Observers at a later epoch would be led to assign the same value for M_0. It is clear in fact that no inference implying the creation of matter could possibly arise in our treatment, since the dynamics is based on counting observations which satisfy Boltzmann's equation.

If in (69) we put $M_0 = c^3 t_0/\gamma_0$, we get

$$\gamma_0 c^3 t_0 m_e^2 m_p = e_0^4. \tag{70}$$

This relation involves only constants, and is independent of choice of t_0, since $e_0^4 \propto t_0^2$, $\gamma_0 \propto t_0$. From this the value of t_0 may be derived We find

$$t_0 = 2{\cdot}15 \times 10^{16}\,\text{sec.} = 0{\cdot}7 \times 10^9\ \text{years},$$

which is of the observed order of magnitude of the present value of t, as it should be. A theoretical derivation of (70), if possible at all, might disclose a factor like π.

SUMMARY

Following the mathematical treatment of the preceding paper, a physical discussion is given in which it is shown that the concept of an electromagnetic field (**E**, **H**) as existing in free space independent of the circumstances of the test-charge used to measure it must be abandoned. Instead it is replaced by the concept of a pair (**E**, **H**) which depend on the velocity **V** of the test-charge at the point concerned used to measure it. Nevertheless "field identities" are found which are satisfied at each event by the (**E**, **H**) as measured there; these identities are explicitly independent of the velocity **V** of the test-charge employed. These are stated as equations (18), (19), (20), (21). Two of them coincide with two of Maxwell's equations, namely those expressing the non-existence of isolated magnetic poles and Faraday's law of electromagnetic induction. The remaining two

are modifications of Maxwell's other two equations for free space. They imply wave-propagation of **E**, **H**, and admit the existence of a vector-potential.

With the abandonment of the concept of a field as existing in free space independent of the velocity of the test-charge used to measure it, the expression of electromagnetic energy as a volume integral must be abandoned also. Instead we have an expression for the electromagnetic energy associated with the different pairs of charges present in the field, calculated mechanically. The sum of this and the mechanical energies of the moving massive particles varies with the time according to a linear function of the accelerations of the charged particles in one another's presence. The latter vanishes in certain well-defined circumstances, or has a mean value zero, and this gives rise to the phenomenon of non-radiating, i.e. "stationary", states. Radiation occurs in certain cases of accelerated motion, but not for periodic systems. The role of the field is then the same as that put forward by Bohr, Kramers and Slater, namely that energy is not located in the field but that the field is the mechanism for conveying energy from one set of charged particles to another.

A unified treatment of gravitational and electromagnetic phenomena is briefly sketched, and the energy formula obtained. A simple Bohr atom is treated on the present formulation of electrodynamics. Dirac's cosmological relation is considered.

REFERENCES

Chalmers and Chalmers 1934 *Phil Mag.* **19**, **436**.
Dirac 1937 *Nature, Lond.*, **139**, **325**.
Milne 1937*a* *Proc. Roy Soc* **A**, **158**, **326**.
— 1937*b* *Proc. Roy. Soc.* **A**, **159**, **171**
— 1937*c* *Proc. Roy. Soc.* **A**, **160**, **1**.
Nernst 1935 *S.B. preuss. Akad. Wiss.* p. **477**.
Whitrow 1936 *Quart. J. Math.* **7**, **271**.

The crystalline structure of steel at fracture

By H. J. Gough, M.B.E., D.Sc., F.R.S. and W. A. Wood, M.Sc.

(*Received* 20 *December* 1937)

[Plates 6—9]

In a previous paper (Gough and Wood 1936) were described the results of a research into the characteristics of deformation and fracture of a mild steel (0·1 % C) under static and fatigue stresses, in which precise methods of X-ray diffraction were used in a systematic study of the changes produced in the crystalline structure by five stressing systems. It was established, for the first time, that failure by static and cyclic stressing was characterized by exactly the same kind of progressive deterioration of the crystalline structure, fracture in all cases being associated with a breakdown—complete or partial—of the crystal grains to a mass of crystallites having a limiting size of between 10^{-4} and 10^{-5} cm. and a completely random orientation. Under static stressing the sequence of changes in structure were successfully studied, while, using cyclic stresses, the separate and combined influences of the range of stress and the superior stress of the cycle were investigated, also the essential differences on the structure between the effects of safe and unsafe ranges of stress were clearly established.

The present paper describes the results of an investigation undertaken in the hope of obtaining a clearer insight into one of the features revealed by the previous research. In addition to a complete fragmentation of the structure into the limiting size of crystallite, it was considered that a state of marked lattice distortion in the fragmented material was also a necessary condition of the fracture stage, but, as stated in the previous paper, the recorded radial broadening of the reflexion spots could not be ascribed with certainty to the presence of distortion in the lattice of the crystallites. It is very difficult to accept the view that the fracture stage is reached as a consequence of fragmented structure alone, for the employment of severe deformations to obtain cold-worked materials possessing increased hardness and increased nominal strength is common practice. These considerations suggested that further light might be thrown on this very important problem by the investigation of a material similar to that used previously but commencing with a structure which had already been brought to a fragmented condition by cold-rolling.

A supply of mild steel, in the normalized condition, was obtained, some of which was cold-worked by rolling so that a reduction in cross-sectional area of 49 % resulted. The structure of this cold-rolled material was found to be entirely fragmented, thus being very suitable for the required purpose, as the characteristics of deformation and fracture would not be confused by further fragmentation and should thus afford direct information on the conditions of the initiation of fracture. Simultaneously with the investigation of the material in this deformed condition, further work has also been carried out on the steel in its normalized state: an improved X-ray technique was devised so that the behaviour of the same identical grains could be followed through successive stages of an experiment.

Throughout the present experiments the fatigue method of testing has been used, as possessing the unique advantage that a progressive fracture can be produced after a great number of applications of the same stress conditions, thus enabling the onset of fracture to be examined at as many stages as desired. Of the many types of cyclic stressing available, choice fell upon that of reversed direct stresses as one by which specimens of the present material can be fractured without undergoing any appreciable change in external shape or dimensions.

With regard to the tests made on the normalized material, while the essential process of deterioration of structure is as established in the previous paper, the improved X-ray technique has established that repeated cycles of a safe range of stress do not, *as* fatigue stresses, produce any appreciable effect on the crystalline structure directly the safe range of stress is exceeded, however, a complete modification of structure results. Another very informative feature established by these tests is that the *rate* of deterioration of structure, due to an unsafe range of stress, decreases as the test progresses until a stable state is approached but not quite reached: the rate of deterioration accelerates during the last stage of the test and fracture results. These structure characteristics, which resemble other fatigue characteristics, suggest that the process of fragmentation is essentially of the nature of a strengthening effect, assisting the material to resist or retard further changes due to the applied stressing system. If a state of equilibrium is actually attained, fracture would be indefinitely postponed.

Although the above are matters of interest, major importance is attached to the results of the examination of the cold-rolled steel. Under safe stress ranges, no change whatever in the structure could be detected. But under unsafe ranges of stress, a most interesting effect was observed. The material initially gives an X-ray spectrum consisting of a continuous ring.

As the test proceeds, a progressive drop occurs in the intensity of the ring relative to the background, but there is no further radial diffusion, which would indicate further break-up of the crystallites, nor any evidence of a state of preferred orientation. The effect is only to be accounted for by the incidence of heavy lattice distortion in the crystallites.

The experiments thus appear to have shown conclusively that the fracture stage of the steel represents the incidence of a certain condition of severe internal stress in the previously fragmented material, thus carrying to a further stage the work described in the previous paper.

The details of the experimental work will now be given.

MATERIAL

The material used for the investigation was a mild steel* which, on analysis, proved to have the following composition: C, 0·12 %; Si, 0·22 %; Mn, 0·62 %; S, 0·008 %; P, 0·018 %, Ni, 0·06 %; Cr (trace). It was thus very similar to, although not of the same batch, as the steel used in the previous research (Gough and Wood 1936). The steel was supplied by the makers in the form of ⅞ in. diameter bar and was stated to have been normalized at 900° C.: a metallurgical examination made at the N.P.L. showed that the structure was typical of a low carbon steel in the hot-rolled condition. Some of the tests to be described were made on the material in the normalized condition as received. To provide cold-worked material, a bar of this normalized steel was cold-rolled so that its diameter was reduced from ⅞ to ⅝ in., giving a reduction in area of 49 %: this reduction was effected in thirteen passes, two through each of five holes and a further three passes through a sixth hole. No subsequent heat treatment was applied to the cold-rolled material which was tested in that condition. These two states of the steel will be referred to as the "normalized" and "cold-rolled" conditions.

STATIC TENSILE PROPERTIES

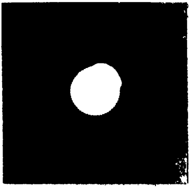
The principal static tensile properties of the steel in each condition were determined by tests made in a Dalby Autographic Recorder: the diagrams obtained are reproduced in fig. 1, while the deduced data are as follows:

(1) *Steel in normalized condition* (specimen 2 A 13, diameter 0·500 in.)

Upper yield stress, 18·0 tons/in.^{2}; lower yield stress, 17·2 tons/in.^{2};

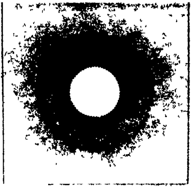
---

* N.P.L. Reference Mark: JPH.



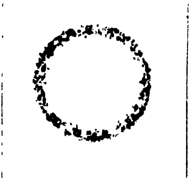
2A11 Initial

Fig. 3.



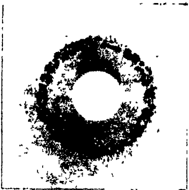
2A11 ±1½ 10^7

Fig. 4.



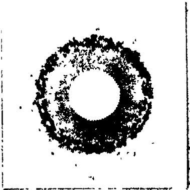
2A12 Initial

Fig. 5



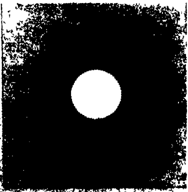
2A12 ±12 10^3

Fig. 6



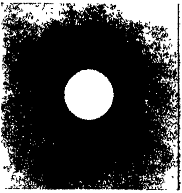
2A12 ±12 10^5

Fig. 7.



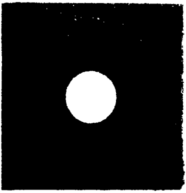
2A12 ±12 10^6

Fig. 8



2A12 ±12 10^7

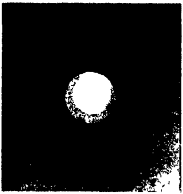
Fig.9.



2A12 ±12 $3\cdot03\times10^7$

Away from Fracture

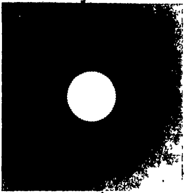
Fig.10.



2A12 ±12 $3\cdot03\times10^7$

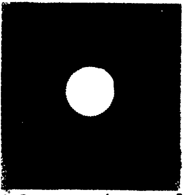
At Fracture

Fig.11.



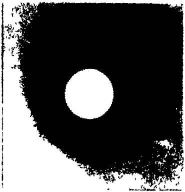
1A8 ±19½ Initial

Fig.12



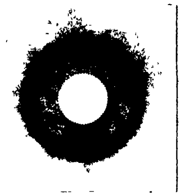
1A8 ±19½ 10^3

Fig.13.



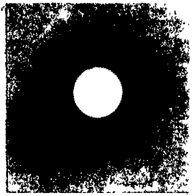
1A8 ±19½ 10^7

Fig.14.



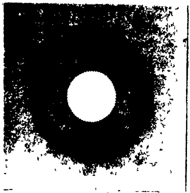
1A10 ±20½ Initial

Fig. 15



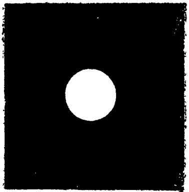
1A10 ±20½ 10^3

Fig. 16.



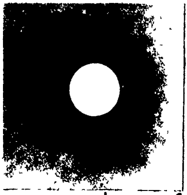
1A10 ±20½ 10^6

Fig. 17



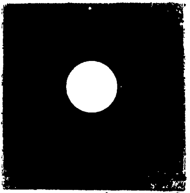
1A10 ±20½ 2×10^6

Fig. 18



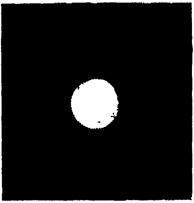
1A10 ±20½ 3×10^6

Fig. 19

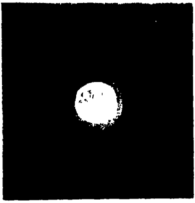


1A10 ±20½ $1{\cdot}14\times10^7$

Fig. 20



Specimen ICI
Broken in static tension
(remote from actual fracture)
Fig. 21



Specimen ICI
Broken in static tension
(at fracture)
Fig. 22

ultimate tensile strength, 28·2 tons/in.^{2}; breaking stress (on final area), 68 tons/in.^{2}; elongation at fracture (on 4 in. gauge length), 33½ %, reduction of area at fracture, 72 %; "cup and cone" type of fracture. Attention is drawn to the drop of stress at the yield point (see fig. 1 *a*), a usual characteristic of mild steel.

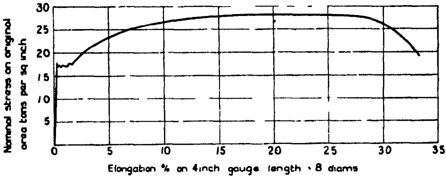


FIG. 1 *a*

(2) *Steel in cold-rolled condition* (specimen 1C1, diameter 0·374 in.)

No yield point: ultimate tensile strength, 56·6 tons/in.^{2}, breaking stress (on final area), 73 tons/in.^{2}; elongation at fracture (on 3 in. gauge length), 5 %, reduction of area at fracture, 51 %, "cup and cone" type of fracture. The absence of a yield point and the form of diagram obtained (see fig. 1 *b*) are characteristic of cold-rolled steel.

Incidentally, data such as are given above are usually interpreted as indicating that the cold-rolling operation has conferred additional strength or hardness above that exhibited during a tensile test made on the material when in the normalized condition, but this has not been demonstrated in the above case. For considering the cold-rolled condition merely as an intermediate stage in the deformation of the original material and making correction for the change in diameter suffered in the cold-rolling, then the ultimate tensile strength and reduction of area at fracture of the cold-rolled steel become 28·9 tons/in.^{2} and 75 %, respectively, as against 28·2 tons/in.^{2} and 72 % for the normalized condition. These figures correspond within the accuracy of the experiment, while the recorded figures of 68 tons/in.^{2} and 73 tons/in.^{2} for breaking strength—which need no correction for change in area—again agree sufficiently closely as to afford no real evidence of strengthening by cold-rolling.

Fatigue test methods

It was decided that, in the present tests, the progressive changes due to fatigue should be complicated by superimposed plastic deformation as little as possible; stress cycles of reversed direct loading (alternating tension and compression), having a numerical average value of zero, were therefore selected for application. As a result, none of the specimens exhibited any measurable change in dimensions after test. the diameter of each was carefully measured, at every stage of test, using accurate optical projection methods. The fatigue tests were carried out in a well-known form of electro-magnetic testing machine at a frequency of 2200 stress cycles per minute circular specimens (see fig. 2 *b*, Gough and Wood, 1936) were employed, the diameters being 0·217 and 0·172 in. for the normalized and cold-rolled conditions, respectively

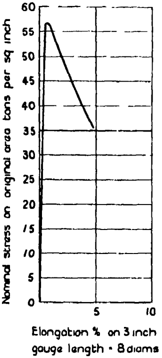


Fig. 1 *b*

Method of X-ray examination

The X-ray technique used was that of the back-reflexion method as in the previous work with, however, two important modifications. First, the

X-ray spectrometer was adapted so that a specimen, examined at stages during its history, could be photographed each time at exactly the same point on the surface. It was possible, as a result, to follow the behaviour in detail of the same set of grains. This represents a definite advance on earlier X-ray technique, especially in the case of normalized material from which the incident beam is diffracted into a series of separate reflexion spots. Any process causing dislocation or rotation, throwing a grain out of the reflecting position, is thus directly emphasized when photographs taken at successive stages are compared. On the other hand, since the number of reflexion spots involved is large, if photographs taken at different stages of a test are identical, spot for spot, the method shows, with a very high degree of sensitivity, that no change has taken place. The second feature utilized an effect first noted in other work on heavily deformed metals. It has been known for some time that the cold working of a metal affects the efficiency of diffraction for X-rays An appropriate wave-length for cold-worked mild steel is the Co-K radiation, which therefore was employed in these experiments.

RESULTS

For each condition of the steel, the procedure adopted was, briefly, as follows. A series of endurance tests was first carried out, each individual test being made without interruption, in order to establish definitely the limiting range of stress for an indefinitely large number of stress cycles. In the majority of cases the specimen was submitted to X-ray examination before and after test only, thus giving information regarding the total changes produced. The limiting ranges thus being established, further specimens of each type were tested at stress ranges slightly greater and, also, slightly less than the limiting range, each test being interrupted at regular (logarithmic) intervals for an X-ray examination in order to make a careful study of the *progressive* changes in structure which occurred. It may be mentioned that previous experience—confirmed by the present tests—had shown that during such interruptions, involving periods of rest, some degree of "recovery" takes place so that the total endurance of such a specimen subjected to such an interrupted test usually exceeds that of a specimen tested without interruption.

The results of the endurance tests which were not interrupted are as stated in Table I, from which it is seen that the limiting fatigue ranges of the metal in the normalized and cold-rolled conditions were clearly defined at ± 11·6 tons/in.^{2} and ± 19·7 tons/in.^{2}, respectively. [These are nominal values reckoned on the cross-sectional areas of the specimens before test. It has been pointed out that the apparent ultimate tensile strength of the cold-rolled material, when corrected for the reduction of area suffered in the cold-rolling operation, agreed very closely with the ultimate tensile strength of the normalized material. Using the same correction, the fatigue strength of the rolled material would be estimated at

$$\pm 11{\cdot}6 \times \frac{1}{0{\cdot}51} = \pm 22{\cdot}7 \text{ tons/in.}^2.$$

The experimental value obtained is ± 19·7 tons/in.^{2} only, suggesting that the intrinsic fatigue strength of the metal has actually suffered *deterioration* during the operation of cold-rolling.]

TABLE I. RESULTS OF FATIGUE TESTS MADE WITHOUT INTERRUPTION

Condition of steel	Reference mark of specimen	Nominal applied range of direct stresses tons/in.^{2}	Total number of cycles endured* (millions of cycles)	Deduced limiting range of stress tons/in.^{2}
Normalized	2A 6	± 13	1·640 B	± 11·6
	2A 7	± 12½	2·172 B	
	2A 8	± 12	4·148 B	
	2A 9	± 11¾	6·118 B	
	2A 10	± 11½	31·012 U	
Cold-rolled	1A 1	± 25	0·330 B	± 19·7
	1A 2	± 24½	0·396 B	
	1A 3	± 23	0·912 B	
	1A 4	± 22	1·598 B	
	1A 5	± 21	0·898 B	
	1A 6	± 20	2·230 B	
	1A 9	± 19½	41·068 U	
	1A 7	± 19	53·576 U	

* B denotes specimen cracked or fractured at end of test. U denotes specimen remaining unbroken at end of test.

The results of the uninterrupted fatigue tests are shown, diagrammatically, in fig. 2.

Turning to the X-ray examination, it is not necessary to describe the results of the examinations made on the specimens whose histories are recorded in Table I, as the features disclosed are fully covered by the characteristics now to be described of those specimens which were examined at progressive stages throughout each test. With the exception of fig. 11 all the X-ray photographs reproduced with this paper were taken at positions on the specimen through which the final fatigue fracture did *not* pass. Photographs taken at the fracture show fragmented structure of which a sufficiently representative number of photographs were reproduced with the previous paper.

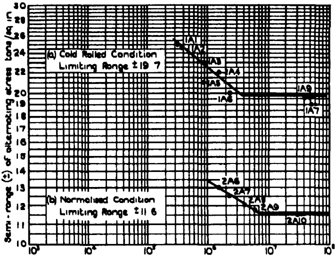


FIG. 2

Normalized condition

Considering, first, the results obtained on the steel in the normalized condition, the application of the improved X-ray technique disclosed features of interest which were additional to those described in the previous paper on very similar material. These may be discussed in relation to the results obtained on the following typical specimens, representing the behaviour of the crystalline structure under repeated cycles of reversed direct stress (*a*) at a range ($\pm 11\frac{1}{2}$ tons/in ^{2}) just less than the fatigue range ($\pm 11{\cdot}7$ tons/in.^{2}), and (*b*) at ranges (± 12 and ± 13 tons/in.^{2}) just above this critical value. The particulars of the specimens used and of the stages at which an X-ray determination was made are summarized in Table II.

Specimen 2 *A*11. This specimen was unbroken after 10^7 cycles of $\pm 11\frac{1}{2}$ tons/in.^{2}, after which the test was discontinued. In the initial condition the specimen consists entirely of large perfect grains giving on the X-ray photograph, as shown in fig. 3, a system of sharp separated reflexion spots. The photographs obtained at each of the five stages of the subsequent history were exactly the same, spot for spot, despite the very large number of reflexions involved. It is only necessary to reproduce the photograph obtained after 10^7 cycles, fig. 4, for comparison with the original structure of fig. 3. This result therefore demonstrates, with a degree of sensitivity not hitherto employed, that repeated cycles of a safe range of stress do not, of themselves, produce any effect whatever on the crystalline structure. in connexion with this statement it will be recalled that, under the present mode of stressing, the specimen exhibits no measurable overall change in dimensions, so that this result is not complicated by the effects of any superposed plastic deformation and this observation applies also to the following specimens.

TABLE II. INTERRUPTED TESTS MADE ON SPECIMENS OF STEEL IN THE NORMALIZED CONDITION

Specimen	Stress range tons/sq. in.	Stages of X-ray examination (cycles)
2A11	$\pm 11\frac{1}{2}$	After 0, 10^3, 10^4, 10^5, 10^6 and 10^7 (unbroken)
2A12*	± 12	After 0, 10^3, 10^4, 10^5, 10^6, 2×10^6, 3×10^6, 4×10^6, 10^7, $3·03 \times 10^7$ (fractured)
2A15*	± 13	After 0, 10^3, 10^4, 10^5, 10^6, 2×10^6, 5×10^6, 5·504 (fractured)

* The total endurances of these specimens are greater than would be predicted from the results given in Table I and fig. 2 which relate to tests carried out without interruption. In our experience, this increased endurance is always a consequence of interruptions and rest periods.

*Specimen 2 A*12. This specimen, tested at ± 12 tons/in.^{2}, represents a very interesting case. The initial photograph, fig. 5, is of the standard type for the normalized material, showing sharply defined reflexion spots. Fig. 6, obtained after 10^3 cycles, however, shows a complete change in the identity of the spots recorded: the replacement of the original spots by fresh ones indicates a modification in structure which could not be detected by earlier technique. Further changes of this nature were also clearly visible after 10^4 and after 10^5 cycles. A further examination made after 10^6 cycles revealed only a small further change and the rate of change in the process of deterioration of structure was obviously slowing up to such an extent that changes in the features of the photographs taken after 10^6, 2×10^6, 3×10^6, 4×10^6 and 10^7 might easily escape attention. Figs. 7 (after 10^5 cycles), 8 (after 10^6 cycles) and 9 (after 10^7 cycles) will sufficiently illustrate these characteristics. there is a distinct although slight difference between figs. 7 and 9, although little between 7 and 8. This specimen, which finally fractured after 3×10^7 cycles, was tested at a range of stress (± 12 tons/in.^{2}) which only slightly exceeded the limiting safe range of stress (± 11·7 tons/in.^{2}): the material is clearly subjected to a critical stress condition. The initial comparatively rapid rate of deterioration, decreases until a stable state is nearly, but not quite, reached. This rate accelerated again during the very last stages of the test as shown by fig. 10, which represents the structure after failure of the specimen at $3{\cdot}03 \times 10^7$ cycles. the structure of fig. 10 is entirely different from that of fig. 9. The fracture of the specimen took place at a position some distance away from the spot to which all these photographs relate. This resumption of activity in the last stages of test is reminiscent of the changes in hysteretic behaviour under fatigue stressing as shown by strain and temperature measurements. The fatigue range of strain or stress should be regarded as marking a dividing line between those conditions under which the structure of the material can or cannot attain a really stable condition· with a very uniform material, the limiting range is clearly marked, with less uniform material, this is not the case and the S/N curve often exhibits some considerable "scatter" of results in this region.

Specimen 2 *A*15. The object of the test made on this specimen also was to trace out the changes occurring under an unsafe range of stress of a value (± 13 tons/in ^{2}) slightly greater than that used in the previous experiment on specimen 2 A 12. The structure at a selected spot was examined before the experiment, after its conclusion (when fracture occurred after $5{\cdot}504 \times 10^6$ cycles) and at six intermediate stages, as recorded in Table II. The X-ray photographs obtained after 10^3, 10^4, 10^5, 10^6 and 2×10^6 cycles also showed clearly that a progressive change was in progress, the changes, as before, occurring at a decreasing rate but not slowing up to anything like the same extent as in specimen 2 A 12 With regard to the last stage of the test immediately leading to fracture and marked by an increase in the rate of modification of the structure, this stage must have been confined in the present instance to the last $0{\cdot}504 \times 10^6$ cycles of the test: although progressive changes were definitely visible after 2×10^6 cycles, the next examination made after 5×10^6 cycles showed very little further change, yet the specimen fractured at a total endurance of $5{\cdot}504 \times 10^6$ cycles. It does not appear to be necessary to reproduce photographs taken from this specimen. The essential features observed were exactly those previously illustrated.

Thus, the present tests made on the normalized steel and using the improved and more sensitive X-ray technique, confirm the previous conclusion that the repeated applications of a range of stress less than the limiting fatigue range do not, in themselves, necessarily produce the slightest change in crystalline structure of the material. In addition, these tests have disclosed that, under an unsafe range of stress, the deterioration of the structure caused by repeated cycles is very marked at the commencement of the test and then proceeds at a diminishing rate until just before actual fracture occurs. The smaller the excess of the applied stress range over the limiting range of stress, the nearer the approach to a state of complete stability under the applied loading.

The actual process of grain fragmentation by the dual processes of grain dislocation and the formation of crystallites is, of course, exactly as described in the previous paper. The new technique, whereby the behaviours of individual grains have been studied, has enabled these characteristics to be identified with the disappearance of some reflexions and the appearance of new reflexions corresponding to the dislocated grains. it has also established definitely that, under some cases of fatigue stressing, the fragmentation is confined to occasional grains.

Cold-rolled condition

Turning now to the fatigue characteristics exhibited by the steel in the cold-rolled state before test, typical specimens of this material were also examined at various stages of test under stress ranges in the neighbourhood of the limiting fatigue range which, it will be recalled, had the value of ± 19·7 tons/in.^{2}. Detailed reference may be made to the results of the typical studies made on the two specimens whose cyclic histories are summarized in Table III.

TABLE III. INTERRUPTED TESTS MADE ON SPECIMENS OF STEEL IN THE COLD-ROLLED CONDITION

Specimen	Stress range tons/in.^{2}	Stages of X-ray examination (cycles)
1A8	± 19½	After 0, 10^3, 10^4, 10^5, 10^6, 10^7, 1 107 × 10^7 (unbroken)
1A10	± 20½	After 0, 10^3, 10^4, 10^5, 10^6, 2 × 10^6, 3 × 10^6, 11 374 × 10^6 (fractured)

Specimen 1*A*8 represents the behaviour of specimens which withstood without fracture the applied cyclic stresses. The initial state gave the diffuse diffraction ring of fig. 12. The crystallites into which the grains are broken down, and which are responsible for the continuous ring, must occupy at least the greater part of the volume, for even if any reflexions from existing large grains are present, they are entirely submerged by the continuous ring. From other work on cold-worked metals it has been pointed out (Wood 1935) that there is a limiting lower size, for a given metal, beyond which the size of the crystallites cannot be reduced by further working. This limit appears to have been reached in the present cold-worked steel.

As with the normalized steel, the examination showed that the application of cyclic stresses having this safe range produced no sign of change of any kind. To establish this fact, it is sufficient to compare the initial state of the material, as shown in fig. 12, with figs. 13 and 14 taken after 10^3 and 10^7 cycles As far as can be seen, they are identical. The cold-worked specimens, therefore, obey the same criterion for safety as the normalized

Specimen 1 *A* 10. This specimen shows the effect of repeated cycles of a range of stress ($\pm 20\frac{1}{2}$ tons/in.^{2}) which exceeds the fatigue range ($\pm 19{\cdot}7$ tons/in.^{2}). The initial state, reproduced in fig. 15, was similar to that of the previous specimen. But under cyclic stressing, progressive changes took place leading to a final state which was very different from the original. After 10^3 cycles, the X-ray photograph, as reproduced in fig 16, showed a marked drop in intensity of the ring relative to the background. There was a further large decrease after 10^6 cycles, to such an extent that, at the last stage before fracture, it was only just possible to pick out the ring from the background. The effect is illustrated in the photographs reproduced in figs. 17–20 inclusive, which show the condition of the structure after 10^6, 2×10^6, 3×10^6 and $1{\cdot}14 \times 10^7$ cycles, respectively. These represent therefore, the progressive changes, leading to fracture, found in the crystalline structure of a material which, prior to the experiments, had been cold-worked sufficiently to prevent any appreciable compensating effects due to further fragmentation of the grains.

An estimation was made from microphotometer measurements of the magnitude of the intensity change. The X-ray tube used had a double window, one of which was used for photographing the test specimens and the other for a standard specimen, so that the efficiency of the tube could be checked. Actually there was no change during the tests. The processing of the negatives was standardized so that the measurements were on the same basis: the measurements obtained are as stated in Table IV, which gives the difference in density* between the peak of the diffraction ring and the background.

TABLE IV. PROGRESSIVE CHANGES IN DENSITY BETWEEN PEAK OF DIFFRACTION RING AND BACKGROUND

No. of cycles ...	0	10^3	10^4	10^5	10^6	2×10^6	3×10^6
Intensity diff. ...	0·21	0·14	0·13	0·13	0·07	0·05	0·03

---

* $D = \log I_0/I$, where D = density, I_0 = intensity of light transmitted through clear film, I = intensity transmitted through point where measurement is taken.

There is thus a final drop of 86 % in the intensity. This is far beyond any experimental error, which would be of the order of a few per cent.

As in the fatigue specimens, the same weakening of the diffraction ring is exhibited by the material when fractured by static tensile forces: figs. 21 and 22 relate to specimen 1C1 and illustrate the effect. Fig. 21 shows the state of the material at the fracture position, where the specimen underwent local reduction of area, while fig. 22 was taken at a position where the specimen had suffered general elongation only.

Apart from the above intensity changes, there were no other changes observed in the diffraction ring. In particular there was no further radial diffusion, which would have accompanied further break-up of the crystallites. Also, there was no departure from the original uniform distribution of intensity round the circumference; no state of preferred orientation of the crystallites had been created. The effect observed is only, therefore, to be accounted for by the incidence of heavy lattice distortion in the crystallites. This distortion appears to be characterized by a "staggering" of the atoms about their normal positions in the crystal lattice, a process which would lead to the observed fall in intensity in a manner analogous to the influence of elevated temperature.

The experiments thus appear to possess considerable significance as affording a direct explanation, supported by physical evidence, of the conditions at fracture of a metal. The stability of the atomic arrangement in the metal is the result of equilibrium between the positive ions and the surrounding electron distribution. Permanent deformation of the metal produces, in the first place, a process of grain dislocation and fragmentation leading to a structure consisting entirely of a mass of crystallites having a limiting size and completely random orientation. But the attainment of this condition is not the criterion of fracture· in fact, although it is doubtful if the metal is really "strengthened" by such cold-working, it can certainly withstand, without plastic deformation, a greater strain or range of strain than in its initial state. There is also little doubt that this process of fragmentation, even in its early stages, has an influence on the stability of the atomic structure as shown, for example, by the well-known effect of cold-work in lowering the recrystallization temperatures of metals: the strain conditions at the boundaries of neighbouring crystallites of different orientation are probably responsible. But the investigation shows that the stage when fracture is imminent is approached only *after* the structure has become completely fragmented. Further cold working does not result in a further reduction in *size* of the crystallite but probably leads to a progressive distortion of the intrinsic structure of the crystallites themselves, the structure as a mass continuing to exhibit random orientation. Whether that is the correct explanation or not of the setting up of severe internal strains, the experiments have clearly demonstrated the progressive deterioration in the regularity of the atomic structure and electron distribution and, therefore, a disturbance of the equilibrium of the structure producing localized regions of weakness at which fracture can be initiated under the action of external forces whose magnitude would appear to be quite inadequate if the material were homogeneous and possessed the full theoretical strength.

ACKNOWLEDGEMENTS

This investigation has been carried out as part of a general research into Elasticity and Fatigue, financed by the Advisory Council of the Department of Scientific and Industrial Research, to whom the authors desire to record their thanks for the research facilities afforded and for permission to publish the results· also to Dr C. H. Desch, F R.S., who kindly undertook the cold-rolling of the steel, and to Mr P. L Thorpe for assistance in the observations.

REFERENCES

Gough, H. J. and Wood, W. A. 1936 *Proc. Roy. Soc.* A, **154**, 510–39.
Wood, W. A. 1935 *Trans. Faraday Soc.* **31**, 1248.

Collective electron ferromagnetism

By Edmund C Stoner, F.R.S.
Reader in Physics at the University of Leeds

(*Received* 29 *December* 1937)

1. Introduction

The essential contribution of Heisenberg to the theory of ferromagnetism was in showing that those effects which had been correlated by means of the formal molecular field hypothesis of Weiss could be interpreted as arising from interchange interaction between electrons in atoms, of the same type as that involved in the formation of homopolar molecules. The Heisenberg method of approach has, however, proved in many ways less convenient in the detailed treatment of ferromagnetism than the method initiated by Bloch for the theory of metallic properties generally, in which possible energy states are derived for electrons treated as waves travelling through the whole crystal The first approximation in this collective electron treatment is that of free electrons, for which the energy is purely kinetic, the number of states per unit energy range then being proportional to the square root of the energy. The effect of the periodic field of the lattice is to modify the distribution of states, giving rise to a series of energy bands, separate or overlapping. Elaborate calculation is necessary to determine the form of these bands with any precision, though in general near the bottom of a band the energy density of states depends on the energy in the same way as for free electrons, but with a different proportionality factor, this holds also near the top of a band, the energy being measured downwards from that limit. The salient characteristics of metals depend on the electrons in unfilled bands. In particular, in the ferromagnetic metals, iron, cobalt and nickel, the ferromagnetism may be attributed to the electrons in the partially filled band corresponding to the *d* electron states in the free atoms. The exchange interaction is such that, at low temperatures, instead of the electrons occupying the lowest states in balanced pairs, there is an excess of electrons with spins pointing in one direction, giving rise to a spontaneous magnetization. The decrease of energy due to the exchange effect with increase in the number of excess parallel spins is accompanied by an increase due to the electrons moving to states of higher energy in the band. The equilibrium magnetization depends on the number of electrons, the form of the band, the magnitude of the exchange interaction, and the temperature, and must be calculated on the basis of Fermi-Dirac statistics. The primary purpose in this paper is the determination of the form of the magnetization temperature curves for bands of the standard type and for a range of values of the exchange interaction energy. Before discussing the particular problem more fully, the relation between the present and some of the previous work will be briefly indicated.

The advantages of a collective electron treatment for ferromagnetism were pointed out some years ago in a paper (Stoner 1933) in which it was shown that such a treatment enabled an immediate interpretation to be given to the non-integral values of the atomic moments of the ferromagnetic metals, and of the variation of the moment with small additions of non-ferromagnetic metals in alloys Little was then known about the form of the electronic energy bands in transition metals, and the treatment was necessarily qualitative. Considerably greater precision in formulation became possible on the basis of a suggestion by Mott (1935), having a general justification, on the form of the energy bands in the ferromagnetic metals. In nickel, in particular, it was suggested that a narrow *d* band was overlapped by a much wider *s* band, and that the top of the Fermi distribution came at a point corresponding to 0·6 electron/atom in the *s* band, with a deficit of the same number in the *d* band, this number corresponding to the observed saturation moment of nickel at low temperatures. (It may be noted that "holes" in bands are to a large extent magnetically equivalent to the same number of electrons in otherwise empty bands, as is more fully discussed elsewhere (Stoner 1936*a*)) This general idea of Mott has been developed successfully in a number of directions (cf. Mott and Jones 1936). A first step in the quantitative treatment of the effect of temperature on the magnetic properties of metals was the determination of the temperature dependence of free electron susceptibility (Stoner 1935, 1936*b*), the results obtained are applicable not only to free electrons but also to electrons in unfilled bands with the same type of energy distribution of states. Series expressions were derived appropriate to high and low temperatures, and values for the intermediate temperature range were found by graphical interpolation. In a subsequent discussion on spin paramagnetism in metals (Stoner 1936*a*) a simple method of taking into account the effect of interchange interaction was described, the method being applicable to the determination of the temperature variation of the susceptibility of a ferromagnetic above the Curie point. Shortly afterwards the treatment was extended to deal with the spontaneous magnetization below the Curie point and the general character of the modifications resulting from the use of Fermi-Dirac, in place of classical, statistics, was determined. It was found, however, that the method previously used, involving series calculations for high and low temperatures, and graphical interpolation, was inadequate to give numerical results of satisfactory precision, the lack of precision being most marked for the temperature range which was often of greatest interest as including the Curie point region. For precise numerical results, an accurate evaluation of a series of the basic Fermi-Dirac functions was indispensable. An extensive table of Fermi-Dirac functions is now available (McDougall and Stoner 1938) and this provides the necessary starting point for much of the computational work involved in the present paper.

In the meantime two papers have appeared by Slater (1936*a*, *b*) on the ferromagnetism of nickel, the first of which deals with the theoretical calculation of the low temperature saturation moment, the second with the temperature variation of the magnetization. An estimate is made of the form of the electronic energy bands, based on extrapolation of calculations for copper by the Wigner-Seitz method, the results confirming the essentials of Mott's suggestion, a theoretical calculation is also made of the magnitude of the exchange interaction. Numerical calculation then shows that the exchange effect is large enough to produce ferromagnetism, and there is fair agreement between the calculated exchange energy and that deduced from the observed Curie temperature. In the second paper numerical calculations are made of the free energy at a series of temperatures as a function of the magnetization, the equilibrium magnetization at each temperature being that for which the free energy is a minimum. The minima are very flat, and in view of the computational difficulties, all that can be said is that there is probably agreement with experiment to within the error of calculation. The method used does not seem adapted to a precise determination of the magnetization as a function of temperature, but the computational difficulties are undoubtedly very formidable when the energy distribution of states in the band is known only in the form of numerical calculations

With the assumption of a "standard" energy distribution of states in a band, precise calculations of the magnetization as a function of the temperature may be made by the methods to be described. Owing to the peculiarities in the forms of the electronic energy bands of particular metals, it cannot be claimed that the results will necessarily be strictly applicable to any actual ferromagnetic. For the results to be applicable with fair approximation, however, it is merely necessary that the band form should approximate to the standard type over part of its range, up to and somewhat beyond the top of the Fermi distribution. As far as can be judged from the graphical representation of the form of the d band in nickel, computed by the Wigner-Seitz method, the approximation is here reasonably close; the energy density of states near the top of the band (the portion here relevant) being roughly proportional to the square root of the energy measured downwards (see Slater 1936*a*, fig. 1, p. 539). It is difficult to say how far the general character of the results is likely to be affected by special peculiarities in the form of bands, but this question opens up such a wide range of possibilities that it is hardly profitable to pursue it, except with reference to more detailed information about the band form for a particular metal. The choice of the standard band form has the formal advantages that precise calculations are possible covering a wide range of variation of the relevant parameters, and that in the limit the results pass over into those obtained on the basis of classical statistics.

In the treatment it is convenient to introduce the parameters ϵ_0, θ', θ and ζ; ϵ_0 being the maximum particle energy at absolute zero (or the energy difference between this and the top of a band, when the "holes" in a band, rather than the electrons, are relevant) in the absence of exchange interaction, $k\theta'$ a measure of the interaction energy, as explained more precisely below, θ the Curie temperature at which the spontaneous magnetization becomes zero, and ζ the relative magnetization, i.e the ratio of the number of excess parallel spins to the total number of potentially effective spins (corresponding in nickel, for example, to the number of holes in the d band). The general problem is to determine ζ as a function of T. The parameter θ' is so chosen that with classical statistics, $\theta' = \theta$; with this statistics, as is well known, ζ/ζ_0 is a unique function of T/θ, and $\zeta_0 = 1$ A single series of calculations is therefore sufficient to determine the ζ, T relations for any value of $k\theta'$. (Such calculations have previously been carried out, Stoner 1931) With Fermi-Dirac statistics an additional parameter, ϵ_0, is involved, and the work necessary to cover an adequate range of values of θ' and ϵ_0 is very much greater It is appropriate to consider ζ as a function of kT/ϵ_0, and to aim at obtaining sets of values of ζ corresponding to different values of $k\theta'/\epsilon_0$. It seems not to be practicable, in general, to obtain these values directly, and the procedure adopted as most convenient has been to calculate $k\theta'/\epsilon_0$ corresponding to different values of ζ (0, 0·1, 0·2, ...) at a series of values of kT/ϵ_0 (0, 0·05, 0·1, ...); from these results, it is then possible, by inverse interpolation, to obtain curves of the more usual form for particular values of $k\theta'/\epsilon_0$. Further, for comparison with the classical ζ/ζ_0, T/θ curve, a series of ζ/ζ_0, T/θ curves may be obtained for a series of values of $k\theta'/\epsilon_0$. The classical curve corresponds to $k\theta'/\epsilon_0 \to \infty$.

The theory is developed so as to cover the paramagnetism above the Curie point as well as the ferromagnetism below, since this must also be considered in any discussion of actual metals. The theoretical formulae are given, when appropriate, in a form suitable for numerical computation. Computational details are briefly described, and the results given in tabular and graphical form The results corresponding to classical statistics are included, as a standard for comparison. A considerably higher degree of precision has been aimed at than might appear necessary for any direct comparison with experiment for two reasons. One is that only a limited number of results can be given within the ranges of two variables, and any interpolation which may be required is hazardous unless the initial values are of adequate precision; the other is that the differences between numbers to be compared are often small, rendering a rather high precision necessary in the numbers themselves if the comparison is to be significant. Further, owing to the character of the functional relations involved, in some parts of the range the attainable precision in the final values obtained, which are of experimental interest, may be less by several powers of ten than that of the values of the basic functions used, so that for these ranges, at least, a much higher precision is essential in the basic computations than that desired in the final results.

2. GENERAL THEORY

The system to be considered is a set of N electrons in a partially filled electronic energy band. The magnetic moment may be calculated from the general formula

$$M = -(\partial F/\partial H)_{T,V}, \tag{2·1}$$

where F is the free energy. With Fermi-Dirac statistics,

$$F = NkT\eta + \Omega, \tag{2·2}$$

where $$\Omega = -kT\sum_s \log[1+\exp\{\eta-(\epsilon_s+\epsilon_s')/kT\}],$$

the summation being taken over all the energy states The energy ϵ_s is taken to include the whole of the energy except that due to the exchange interaction which involves dependence on the magnetization as a whole, and that due to an external field. The corresponding terms make up ϵ_s' which may be written

$$\epsilon_s' = \epsilon_1 + \epsilon_2.$$

The effect of exchange interaction is to introduce a term proportional to the square of the magnetization (for a fuller theoretical discussion, reference may be made to Slater 1936*a*; see also Stoner 1936*a*), so that the corresponding energy per unit volume E_v, may be expressed as

$$E_v = \tfrac{1}{2}\alpha I^2.$$

Since $$\partial E_v/\partial I = -\alpha I,$$

the energy of an electron with spin antiparallel and parallel to I is given by

$$\epsilon_1 = \pm\mu\alpha I.$$

Let M be the resultant magnetic moment of the N electrons, ζ the relative magnetization, and n_v the number of electrons per unit volume. Then,

$$\zeta = M/N\mu, \tag{2·3}$$

$$\epsilon_1 = \pm\mu^2\alpha n_v\zeta.$$

This will be written $$\epsilon_1 = \pm k\theta'\zeta. \tag{2·4}$$

Thus $k\theta'$, as used here, is a measure of the exchange interaction energy, and is such that $2k\theta'\zeta$ is the difference in energy for the spin parallel and antiparallel to the magnetization, while the difference in energy *per* electron in the completely magnetized state ($\zeta = 1$) and the demagnetized state ($\zeta = 0$) is $\frac{1}{2}k\theta'$. (It will be shown below that $\theta' > \theta$, where θ is the Curie temperature except in the classical limit, when $\theta' = \theta$, $k\theta'$ will be referred to as the interaction energy coefficient, that part only of the interaction energy which depends on number of excess parallel spins being understood.) Considering now the effect of the external field, the electron may be treated as though the spin is orientated parallel or antiparallel to the resultant of the external field and the effective internal field due to exchange interaction, so that

$$\epsilon_s' = \epsilon_1 + \epsilon_2 = \pm(k\theta'\zeta + \mu H).$$

Let $\nu(\epsilon)$ be the number of states per unit energy range for one direction of spin; then

$$\Omega = -kT\int_0^\infty \nu(\epsilon)\log[1+\exp\{\eta-(\epsilon-k\theta'\zeta-\mu H)/kT\}]\,d\epsilon$$
$$-kT\int_0^\infty \nu(\epsilon)\log[1+\exp\{\eta-(\epsilon+k\theta'\zeta+\mu H)/kT\}]\,d\epsilon. \tag{2·5}$$

It is convenient to introduce the abbreviations

$$x = \epsilon/kT;\ \beta = k\theta'\zeta/kT = (\theta'/T)\zeta,\ \beta' = \mu H/kT. \tag{2·6}$$

For bands of the standard form,

$$\nu(\epsilon) = a\epsilon^{\frac{1}{2}},$$

and with ϵ_0 as the maximum electron energy when the electrons are in balanced pairs in the lowest energy states,

$$N = 2\int_0^{\epsilon_0} a\epsilon^{\frac{1}{2}}\,d\epsilon = \tfrac{4}{3}a\epsilon_0^{\frac{3}{2}},$$

giving

$$\nu(\epsilon) = \tfrac{3}{4}(N/\epsilon_0^{\frac{3}{2}})\,\epsilon^{\frac{1}{2}}. \tag{2·7}$$

Substituting (2·6) and (2·7) in (2·5),

$$\Omega = -\tfrac{3}{4}NkT(kT/\epsilon_0)^{\frac{3}{2}}\int_0^\infty x^{\frac{1}{2}}[\log\{1+\exp(\eta-x+\beta+\beta')\} \\ \times\{1+\exp(\eta-x-\beta-\beta')\}]\,dx.$$

This formula is simplified by integrating by parts, since

$$\int_0^\infty x^{\frac{1}{2}}\log\{1+\exp(\eta-x)\}\,dx = \frac{2}{3}\int_0^\infty \frac{x^{\frac{3}{2}}\,dx}{e^{x-\eta}+1} = \tfrac{2}{3}F_{\frac{3}{2}}(\eta).$$

Thus,

$$\Omega = -\tfrac{1}{2}NkT(kT/\epsilon_0)^{\frac{3}{2}}[F_{\frac{3}{2}}(\eta+\beta+\beta')+F_{\frac{3}{2}}(\eta-\beta-\beta')] \tag{2·8}$$

Making use of the relation

$$\frac{d}{d\eta}F_{\frac{3}{2}}(\eta) = \tfrac{3}{2}F_{\frac{1}{2}}(\eta) = \frac{3}{2}\int_0^\infty \frac{x^{\frac{1}{2}}\,dx}{e^{x-\eta}+1}, \tag{2·9}$$

expressions are obtained for M and N from the formulae

$$M = -(\partial F/\partial H)_{T,V} = -(\partial\Omega/\partial H)_{\eta,T,V},$$

and

$$N = -(1/kT)(\partial\Omega/\partial\eta),$$

in the form

$$M = \tfrac{3}{4}N\mu(kT/\epsilon_0)^{\frac{3}{2}}[F_{\frac{1}{2}}(\eta+\beta+\beta')-F_{\frac{1}{2}}(\eta-\beta-\beta')], \tag{2·10}$$

$$N = \tfrac{3}{4}N(kT/\epsilon_0)^{\frac{3}{2}}[F_{\frac{1}{2}}(\eta+\beta+\beta')+F_{\frac{1}{2}}(\eta-\beta-\beta')]. \tag{2·11}$$

These are the two fundamental equations of the present treatment, from which more convenient forms appropriate to the various special cases may be derived, as will be shown. The function $F_{\frac{1}{2}}(\eta)$ is one of the Fermi-Dirac functions which have been tabulated for a wide range of values of the argument (McDougall and Stoner 1938), while outside the range covered by the tables it may be accurately computed by using a few terms of series, which have also been given. Since the other functions are not further used in this paper, the subscript will be omitted, and also, where there is no ambiguity, the argument, i.e.

$$F_{\frac{1}{2}}(\eta) = F(\eta) = F. \tag{2·12}$$

Series expansions

The series expansions required are generally based on the two series expansions for $F(\eta)$, one appropriate for $\eta < 0$ (corresponding approximately to $kT/\epsilon_0 > 1$) and the other (as asymptotic series) for $\eta \gg 1$ (approximately $kT/\epsilon_0 \ll 1$).

For $\eta < 0$,
$$F(\eta) = \frac{\sqrt{\pi}}{2}\sum_{r=1}^{\infty} b_r e^{r\eta}, \tag{2·13}$$

where
$$b_r = (-)^{r-1}/r^{\frac{3}{2}}.$$

For $\eta \gg 1$,
$$F(\eta) = \tfrac{2}{3}\eta^{\frac{3}{2}}\left\{1 + \sum_{r=1}^{n} a_{2r}\eta^{-2r}\right\}, \tag{2·14}$$

where
$$a_{2r} = 2c_{2r}\left(\tfrac{3}{2}\right)\left(\tfrac{1}{2}\right)\left(-\tfrac{1}{2}\right)\,..\left(\tfrac{5}{2} - 2r\right),$$

and
$$c_{2r} = \sum_{s=1}^{\infty}(-)^{s-1}s^{-2r}$$

It may be noted that
$$c_2 = \pi^2/12, \quad c_4 = 7\pi^4/720,$$

as these forms often enable the appearance of formulae to be simplified.

The series (2·13) and (2·14) have previously been fully considered, a discussion having been included of the number of terms to be taken in (2·14) to obtain the best approximation (McDougall and Stoner 1938), and the methods for proceeding from series in η to series in (kT/ϵ_0) have been described (Stoner 1935, 1936*a*, *b*), so that it will usually be sufficient in the sequel to give the final series expressions without details of their derivation. As extensive use is made of the inverse series derived from (2·13) and (2·14) it may, however, be well to give these

For $\eta < 0$, writing $y = \dfrac{2}{\sqrt{\pi}}F(\eta)$,

$$e^{\eta} = y\left\{1 + \sum_{r=1}^{\infty} b'_r y^r\right\}, \tag{2·15}$$

where
$$b_1' = \frac{\sqrt{2}}{4} \qquad = 0{\cdot}353\ 553\ 39,$$
$$b_2' = \frac{1}{4} - \frac{\sqrt{3}}{9} \qquad = 0{\cdot}057\ 549\ 91,$$
$$b_3' = \frac{1}{8} + \frac{5\sqrt{2}}{32} - \frac{5\sqrt{6}}{36} = 0{\cdot}005\ 763\ 96.$$

For $\eta \gg 1$, writing $y = \{\frac{3}{2}F(\eta)\}^{\frac{2}{3}}$,
$$\eta = y\left\{1 - \sum_{r=1}^{n} a_{2r}' y^{-2r}\right\},$$
where
$$a_2' = c_2 \qquad = \frac{\pi^2}{12} \qquad = 0{\cdot}822\ 467\ 03,$$
$$a_4' = \tfrac{3}{4}c_2^2 + \tfrac{3}{4}c_4 \qquad = \frac{\pi^4}{80} \qquad = 1{\cdot}217\ 613\ 64,$$
$$a_6' = \tfrac{7}{8}c_2^3 + \tfrac{21}{8}c_2c_4 + \tfrac{105}{16}c_6 = \frac{247\pi^6}{25920} = 9{\cdot}161\ 386\ 22.$$

3. Paramagnetism

(i) *Without interaction*

When the exchange interaction effect is zero ($\beta = 0$) the relative magnetization is given by
$$\zeta = \frac{M}{N\mu} = \frac{F(\eta+\beta') - F(\eta-\beta')}{F(\eta+\beta') + F(\eta-\beta')}, \tag{3·1}$$
where
$$\beta' = \mu H/kT.$$
In the limit $(\epsilon_0/kT) \to 0$, $F(\eta) \to \frac{1}{2}\pi^{\frac{1}{2}}e^{\eta}$, and the classical result is obtained:
$$\zeta = \tanh\beta'. \tag{3·2}$$
Provided that β' is small, so that only the first power in H need be considered,
$$\zeta = \beta'(F'/F), \tag{3·3}$$
giving in the classical limit (for which $F'/F \to 1$), the well-known result
$$M/H = N\mu^2/kT. \tag{3·4}$$
It is usual in dealing with experimental results on paramagnetism to consider the inverse of the susceptibility as a function of temperature. In presenting the theoretical results with Fermi-Dirac statistics, it is con-

venient to consider kT/ϵ_0 as the variable, rather than T, since F and F' are functions of kT/ϵ_0, and to express the relation (3·3) in the form

$$\frac{1}{\zeta}\cdot\frac{\mu H}{\epsilon_0} = \frac{kT}{\epsilon_0}\cdot\frac{F}{F'}, \tag{3·5}$$

the quantity on the left being proportional to the inverse of the susceptibility. The theoretical curves corresponding to this relation have been approximately computed (Stoner 1935, 1936*b*) using the methods indicated in § 1.

(ii) *With interaction*

Provided that the magnetization is proportional to the external field, β and β' may both be treated as small, leading, in place of (3·3), to

$$\zeta = (\beta+\beta')\ (F'/F), \tag{3·6}$$

where $\beta = (\theta'/T)\,\zeta$, so that

$$\zeta = \frac{\mu H}{kT}\cdot\frac{F'}{F}\bigg/\left\{1-\frac{\theta'}{T}\cdot\frac{F'}{F}\right\}. \tag{3·7}$$

The relation for the inverse of the susceptibility, corresponding to (3·5), is

$$\frac{1}{\zeta}\cdot\frac{\mu H}{\epsilon_0} = \frac{kT}{\epsilon_0}\cdot\frac{F}{F'} - \frac{k\theta'}{\epsilon_0}. \tag{3·8}$$

This equation expresses in convenient form the conclusion which was previously drawn (Stoner 1936*a*) that the effect of positive interchange interaction is to decrease by a constant amount the inverse of the susceptibility as calculated with neglect of interaction. From the computed values of the susceptibility without interaction, those for the susceptibility above the Curie point can therefore be immediately obtained for any value of $k\theta'/\epsilon_0$ It may be noted that in the classical limit, (3·8) is equivalent to the Weiss law. The Curie temperature, θ, is the temperature at which, in this approximation, the ratio of field to magnetization becomes zero, so that, from (3·8)

$$\theta'/\theta = F/F'. \tag{3·9}$$

It may be noted that the dependence of θ' on θ is the same (apart from constant factors) as that of the inverse of the susceptibility, without interaction, on T; the relation with (3·5) is brought out by writing (3·9) in the form

$$\frac{k\theta'}{\epsilon_0} = \frac{kT}{\epsilon_0}\cdot\frac{F}{F'}. \tag{3·10}$$

4. FERROMAGNETISM. GENERAL EQUATIONS

The equations determining the spontaneous magnetization in the absence of an external field ($\beta' = 0$) will now be considered. These are (2·10) and (2·11), which may be written in the form

$$F(\eta+\beta)+F(\eta-\beta) = \frac{4}{3}\left(\frac{\epsilon_0}{kT}\right)^{3/2}, \tag{4·1}$$

$$F(\eta+\beta)-F(\eta-\beta) = \frac{4}{3}\left(\frac{\epsilon_0}{kT}\right)^{3/2}\zeta, \tag{4·2}$$

and the equation defining β,

$$\beta = (\theta'/T)\,\zeta. \tag{4·3}$$

As a preliminary to an account of the method of dealing with these equations, it will perhaps be useful to refer briefly to the methods applicable in the much simpler classical problem. In the classical limit ($\epsilon_0/kT \to 0$) the two equations (4·1) and (4·2) reduce to the single equation

$$\zeta = \tanh\beta, \tag{4·4}$$

the simultaneous equations (4·3) and (4·4) then determining the magnetization. It is readily shown (see below) that, in this limit, $\theta' = \theta$, and it is possible to determine ζ for a given value of T/θ by seeking the value of β satisfying

$$\tanh\beta/\beta = T/\theta, \tag{4·5}$$

and from this finding ζ from (4·3). The numerical results may be given as a table of values of ζ at equal intervals of T/θ, generally the most useful form. It is, however, more convenient to determine θ/T for a given value of ζ directly from the equation

$$\frac{\theta}{T} = \frac{\tanh^{-1}\zeta}{\zeta} = \frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta}. \tag{4·6}$$

This method would ordinarily be applied to give values of T/θ at equal intervals of ζ, a form which is less familiar and which has some disadvantages. The first method, however, is virtually impracticable in the Fermi-Dirac problem, and it is the analogue of the second which has been applied.

The ultimate aim with the equations (4·1), (4·2) and (4·3) is to obtain ζ as a function of kT/ϵ_0 for a series of given values of $k\theta'/\epsilon_0$. It is a simple matter to calculate the values of ζ, kT/ϵ_0 and $k\theta'/\epsilon_0$ corresponding to a given pair of values of η and β; but to obtain in this way an adequate series of pairs of values of ζ and kT/ϵ_0 corresponding to a given $k\theta'/\epsilon_0$ would involve an amount of labour in computation and interpolation which would be prohibitive unless only rough values were required. The method finally adopted was to make the primary evaluations those of $k\theta'/\epsilon_0$ for given values of kT/ϵ_0 and ζ. From (4·1) and (4·2),

$$F(\eta+\beta) = \frac{2}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}(1+\zeta), \tag{4·7}$$

$$F(\eta-\beta) = \frac{2}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}(1-\zeta) \tag{4·8}$$

From these equations $(\eta+\beta)$ and $(\eta-\beta)$ may be found by inverse interpolation from the $F(\eta)$ tables or by the series (2·15) and (2·16) outside the range of the tables, for given values of kT/ϵ_0 and ζ Hence η and β may be obtained, and from β, using (4·3), the corresponding value of θ'/T and so of $k\theta'/\epsilon_0$. From a table of values of $k\theta'/\epsilon_0$ at appropriate intervals of kT/ϵ_0 for a given ζ, the value of kT/ϵ_0 for that ζ for any required value of $k\theta'/\epsilon_0$ may then be found by inverse interpolation. The above outlines the method which is generally applicable. For certain parts of the range, however, the computational work may be greatly reduced by making use of relations obtainable in the form of series expansions, which are of additional value for checking purposes Apart from their use as aids to computation, these relations are of value in bringing out some of the essential characteristics of the numerical results; more particularly in showing the character of the change in the dependence of magnetization on temperature resulting from the substitution of Fermi-Dirac for classical statistics. These relations are developed in the next section.

5 FERROMAGNETISM SERIES EXPANSIONS

(i) *The Curie temperature*

The Curie temperature, θ, is the temperature at which the spontaneous magnetization becomes zero. For ζ (and β) small, the left-hand sides of equations (4·2) and (4·3) may be expanded in a Taylor series, and by division

$$\zeta = \{\beta F' + (\beta^3/6)F''' \ldots\}/\{F + (\beta^2/2)F'' \ldots\}.$$

In the limit $\zeta \to 0$ (involving $\beta \to 0$),

$$\zeta = \beta(F'/F) = (\theta'/T)(F'/F)\zeta,$$

the temperature T for which this equation is satisfied being the Curie temperature, θ; so that

$$\theta'/\theta = F/F'. \tag{5·1}$$

This derivation is an alternative to that given above (see equations (3·9) and (3·10)) based on a consideration of the equation for paramagnetism above the Curie point.

The series forms of the relation (5·1) for high and low temperatures are readily obtained.

For $k\theta/\epsilon_0 > 1$, using (2·13),

$$\theta'/\theta = 1 + \sum_{r=1}^{\infty} b_r'' y^r, \tag{5·2}$$

where

$$y = \frac{4}{3\sqrt{\pi}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{3}{2}},$$

$$b_1'' = \frac{\sqrt{2}}{4}, \quad b_2'' = -\left\{\frac{2\sqrt{3}}{9} - \frac{3}{8}\right\}, \quad b_3'' = \left\{\frac{3}{8} + \frac{5\sqrt{2}}{16} - \frac{\sqrt{6}}{3}\right\}.$$

This series expression shows that $\theta'/\theta \to 1$ for $\epsilon_0/k\theta \to 0$, and that the ratio increases as $k\theta/\epsilon_0$ decreases (Numerical values of related coefficients more convenient for computation are given below, equation (5·6).)

For $k\theta/\epsilon_0 \ll 1$, using (2·14),

$$\left.\begin{aligned} \frac{k\theta'}{\epsilon_0} &= \frac{2}{3}\left[1 + \frac{\pi^2}{12}\left(\frac{k\theta}{\epsilon_0}\right)^2 + \frac{3\pi^4}{80}\left(\frac{k\theta}{\epsilon_0}\right)^4 \dots\right], \\ &= \frac{2}{3} + a_2''\left(\frac{k\theta}{\epsilon_0}\right)^2 + a_4''\left(\frac{k\theta}{\epsilon_0}\right)^4 \dots, \end{aligned}\right\} \tag{5·3}$$

where $a_2'' = 0·548\,311$, $a_4'' = 2·435\,227$. This expression is derived from the asymptotic series (2·14), and the $k\theta/\epsilon_0$ range for which it may be usefully applied is not appreciably extended by the inclusion of further terms. The relation shows at once that there is a lower limit to $k\theta'/\epsilon_0$ for the occurrence of spontaneous magnetization at any temperature For ferromagnetism to occur at all (i.e. for $k\theta/\epsilon_0 > 0$) a necessary condition is

$$k\theta'/\epsilon_0 > \tfrac{2}{3} \tag{5·4}$$

(ii) $k\theta'/\epsilon_0$ *as a function of* kT/ϵ_0 *and* ζ

The relation between $k\theta'/\epsilon_0$ and $k\theta/\epsilon_0$ (i.e kT/ϵ_0 for $\zeta = 0$) is a special case of the more general relations between $k\theta/\epsilon_0$ and kT/ϵ_0 for any value of ζ. The procedure in obtaining the more general series expansions is similar, the starting point being equations (4·7) and (4·8). From these, inverse series are obtained for $(\eta+\beta)$ and $(\eta-\beta)$ and by subtraction a series for β and hence for θ'/T (since $\beta=(\theta'/T)\zeta$). The series appropriate for $kT/\epsilon_0>1$ will first be considered. For $\eta+\beta<0$, from (2·15),

$$e^{\eta+\beta}=y_1\left\{1+\sum_{r=1}^{\infty}b'_r\,y_1^r\right\},$$

where the coefficients are those already given, and

$$y_1=\frac{4}{3\sqrt{\pi}}\left(\frac{\epsilon_0}{kT}\right)^{3/2}(1+\zeta)=y(1+\zeta)$$

Similarly
$$e^{\eta-\beta}=y_2\left\{1+\sum_{r=1}^{\infty}b'_r\,y_2^r\right\},$$

where
$$y_2=\frac{4}{3\sqrt{\pi}}\left(\frac{\epsilon_0}{kT}\right)^{3/2}(1-\zeta)=y(1-\zeta)$$

Taking logarithms, and subtracting,

$$2\beta=2\left(\frac{\theta'}{T}\right)\zeta=\log\frac{1+\zeta}{1-\zeta}+\log\frac{1+\Sigma b'_r y_1^r}{1+\Sigma b'_r y_2^r}.$$

Expanding the second term, and dividing by 2,

$$\begin{aligned}\frac{\theta'}{T}&=\frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta}+b'_1y+(2b'_2-b_1'^2)\,y^2\\&\qquad+(3b'_3-3b'_2b'_1+b_1'^3)\,y^3\dots\\&=\frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta}+b''_1y+b''_2y^2+b''_3(1+\zeta^2/3)\,y^3\quad,\end{aligned}\tag{5·5}$$

where the coefficients b''_r are those already given in (5·2) for the series for θ'/θ. On expanding the first term,

$$\frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta}=1+\frac{\zeta^2}{3}+\frac{\zeta^4}{5}\dots,$$

it becomes clear that (5·5) reduces to (5·2) for $\zeta\to0$. For computational purposes it is convenient to express (5·5) as a series in ϵ_0/kT:

$$\frac{\theta'}{T}=\frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta}+b'''_1\left(\frac{\epsilon_0}{kT}\right)^{3/2}+b'''_2\left(\frac{\epsilon_0}{kT}\right)^{3}+b'''_3(1+\zeta^2/3)\left(\frac{\epsilon_0}{kT}\right)^{9/2}\dots,\tag{5·6}$$

where
$$b'''_r=\left(\frac{4}{3\sqrt{\pi}}\right)^r b''_r,$$

giving
$$b'''_1=+0·265\,961\,52,$$
$$b'''_2=-0·005\,602\,36,$$
$$b'''_3=+0·000\,189\,50.$$

The first term in (5·6) corresponds to the classical result (4·6). The most interesting point about the expression (5·6) is that it shows that, although θ'/T may differ greatly from the classical value, unity, for $T = \theta$, the increase in θ'/T with ζ, for T constant, approximates closely to the classical increase (for $\epsilon_0/kT \to 0$) provided that ζ is not too large:

$$\frac{1}{T}(\theta'_\zeta - \theta'_0) = \frac{1}{T}(\theta_\zeta - \theta_0)_{\epsilon_0/kT\to 0} + \tfrac{1}{3}b''_3\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}\zeta^2. \quad (5\cdot7)$$

This result greatly facilitates computation for the larger values of kT/ϵ_0. (The coefficient of ζ^2 ranges from 0·000 063 1 for $kT/\epsilon_0 = 1{\cdot}0$ to 0·000 002 8 for $kT/\epsilon_0 = 2{\cdot}0$, the precision obtainable by this method for $kT/\epsilon_0 > 1{\cdot}0$ is comparable, or higher, than that obtainable by the general computational method, except possibly for $\zeta > 0{\cdot}9$.)

A similar procedure may be followed for $kT/\epsilon_0 \ll 1$, making use of the asymptotic series expansion (2·16). The following expression is obtained.

$$2\beta = 2\left(\frac{\theta'}{T}\right)\zeta = (z_1 - z_2) - \frac{\pi^2}{12}(z_1^{-1} - z_2^{-1}) - \frac{\pi^4}{80}(z_1^{-3} - z_2^{-3}) \ \ldots,$$

where $\quad z_1 = (\epsilon_0/kT)(1+\zeta)^{\frac{2}{3}}, \quad z_2 = (\epsilon_0/kT)(1-\zeta)^{\frac{2}{3}}.$

For $\zeta \ll 1$, neglecting terms beyond those in ζ^2 and $(kT/\epsilon_0)^4$,

$$\frac{k\theta'}{\epsilon_0} = \frac{2}{3}\left[\left(1 + \frac{2}{27}\zeta^2\right) + \frac{\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2\left(1 + \frac{20}{27}\zeta^2\right) + \frac{3\pi^4}{80}\left(\frac{kT}{\epsilon_0}\right)^4(1 + 2\zeta^2)\right], \quad (5\cdot8)$$

a formula agreeing, for $\zeta \to 0$, with that derived for $k\theta'/\epsilon_0$ in terms of $k\theta/\epsilon_0$ (5·3). It is of interest to consider the form of the expression appropriate for $kT/\epsilon_0 \to 0$ Provided that $kT/\epsilon_0 \ll (1-\zeta)^{\frac{2}{3}}$, retaining terms up to that in $(kT/\epsilon_0)^2$,

$$\frac{k\theta'}{\epsilon_0} = \frac{1}{2\zeta}\{(1+\zeta)^{\frac{2}{3}} - (1-\zeta)^{\frac{2}{3}}\}\left[1 + \frac{\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2(1-\zeta^2)^{-\frac{4}{3}}\right] \quad (5\cdot9)$$

Proceeding to the limit, this expression shows that at absolute zero ζ may have a maximum value less than 1 unless $k\theta'/\epsilon_0$ exceeds a certain critical value. Denoting the value of ζ at absolute zero by ζ_0,

$$\frac{k\theta'}{\epsilon_0} = \frac{1}{2\zeta_0}\{(1+\zeta_0)^{\frac{2}{3}} - (1-\zeta_0)^{\frac{2}{3}}\}, \quad (5\cdot10)$$

giving $\quad \zeta_0 < 1 \text{ for } k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}.$

Thus the value of ζ_0 ranges from 0 to 1 as $k\theta'/\epsilon_0$ ranges from $\frac{2}{3}$ to $2^{-\frac{1}{3}}$ (= 0·793 701).

Binomial expansion of (5·10) gives

$$\frac{k\theta'}{\epsilon_0} = \frac{2}{3}\left(1 + \frac{2}{27}\zeta_0^2 + \frac{7}{243}\zeta_0^4 \ldots\right). \tag{5·11}$$

Comparison with (5·3) then yields the approximate relation

$$\left(\frac{k\theta}{\epsilon_0}\right)^2 = \frac{2}{27}\cdot\frac{12}{\pi^2}\zeta_0^2\left(1 - \frac{1}{90}\zeta_0^2\right),$$

or

$$\frac{k\theta}{\epsilon_0} = \frac{2\sqrt{2}}{3\pi}\zeta_0\left(1 - \frac{1}{180}\zeta_0^2\right), \tag{5·12}$$

showing that the Curie temperature is closely proportional to the maximum magnetization, for small values of ζ_0 (and so of $k\theta/\epsilon_0$). The approximate relation may be expressed numerically by

$$k\theta/\epsilon_0 = 0{\cdot}300\,105\,\zeta_0. \tag{5·13}$$

(iii) *Magnetization near the Curie point*

Even in the classical problem it is not possible to obtain convenient expressions giving ζ as an explicit function of T/θ for the whole of the range, but useful expressions may be given for $T/\theta \to 1$. The analogues of these appropriate for $k\theta/\epsilon_0 \gg 1$ and $k\theta/\epsilon_0 \ll 1$ will now be derived. The classical method is to express ζ first as a power series in β, this method may be followed with Fermi-Dirac statistics, but as it involves subsequent elimination of η, it is simpler to proceed directly from the expressions already obtained for $k\theta'/\epsilon_0$ as a function of kT/ϵ_0 and ζ. As the later terms in the series expansions are unwieldy, and are not required for computation, the leading terms only will be considered.

For $k\theta/\epsilon_0 > 1$, $\zeta \ll 1$, the use of equations (5·2) and (5·6), retaining terms up to those in y or $(\epsilon_0/kT)^{\frac{1}{2}}$ and ζ^2, yields

$$\frac{T}{\theta} = 1 - \frac{\zeta^2}{3} + \frac{1}{3}\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left\{\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{1}{2}} - \left(\frac{\epsilon_0}{kT}\right)^{\frac{1}{2}}\right\},$$

or, for $\{1-(T/\theta)\}\to 0$,

$$\zeta^2 = 3\left(1 - \frac{T}{\theta}\right)\left\{1 - \frac{1}{3\sqrt{(2\pi)}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{1}{2}}\right\}, \tag{5·14}$$

giving

$$-\left[\frac{d\zeta^2}{d(T/\theta)}\right]_{T/\theta\to 1} = 3\left\{1 - \frac{1}{3\sqrt{(2\pi)}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{1}{2}}\right\}. \tag{5·15}$$

For computational purposes it may be noted that

$$\frac{1}{3\sqrt{(2\pi)}} = 0{\cdot}132\,981.$$

The classical result is obtained for the limit $\epsilon_0/k\theta \to 0$; in general the slope of the ζ^2, (T/θ) curve at the Curie point is less than that given by classical statistics.

For $k\theta/\epsilon_0 \ll 1$, $\zeta \to 0$, the result obtained, using (5·3) and (5·8), is

$$\zeta^2 = \frac{27}{2} \cdot \frac{\pi^2}{12}\left(\frac{k\theta}{\epsilon_0}\right)^2 \left\{1 - \left(\frac{T}{\theta}\right)^2\right\}, \tag{5·16}$$

giving

$$-\left[\frac{d\zeta^2}{d(T/\theta)}\right]_{T/\theta \to 1} = \frac{9\pi^2}{4}\left(\frac{k\theta}{\epsilon_0}\right)^2$$

$$= 22{\cdot}206\,610\ (k\theta/\epsilon_0)^2.$$

For the low values of $k\theta/\epsilon_0$ for which this result applies the slope of the ζ^2, (T/θ) curve will be much less than that corresponding to (5·15). In this region ζ_0 is small compared with unity and it is legitimate to substitute for $k\theta/\epsilon_0$ from (5·12), leading to an alternative approximate expression,

$$\zeta^2 = \zeta_0^2\{1 - (T/\theta)^2\}, \tag{5·17}$$

which gives

$$-\left[\frac{d(\zeta/\zeta_0)^2}{d(T/\theta)}\right]_{T/\theta \to 1,\ \zeta_0 \to 0} = 2$$

The numerical results given later show that (5·17) holds, as a formula for T/θ, to an accuracy of within 1 % up to $\zeta_0 = 0{\cdot}6$ for $\zeta/\zeta_0 < 0{\cdot}6$ (see Table III, fig. 3). In the range for which (5·15) is applicable, $\zeta_0 = 1$. The slope of the $(\zeta/\zeta_0)^2$, (T/θ) curve for $T/\theta \to 1$ thus changes from 3 to 2 as $k\theta/\epsilon_0$ changes from high to low values. The change, however, is not monotonic, the numerical results showing that values somewhat lower than 2 occur in the intermediate range

(iv) *Approach to saturation*

It is hardly possible to obtain convenient series expressions for the approach to saturation at absolute zero having more than a very limited range of applicability; for the values of the temperature and magnetization are such that, in general, entirely different forms are appropriate for the approximate expression of the basic functions $F(\eta+\beta)$ and $F(\eta-\beta)$, and the approximation given by simple expressions is not usually very close.

Provided that $\zeta_0 = 1$, it is clear from (4·7) and (4·8) for $\zeta \to \zeta_0$,

$$\{F(\eta-\beta)\}/\{F(\eta+\beta)\} \to 0.$$

Using (4·1) and (4·2) an approximate expression for the magnetization is

$$\zeta = 1 - 2\frac{F(\eta-\beta)}{F(\eta+\beta)}, \tag{5·18}$$

which, for $\zeta \to 1$, becomes

$$\zeta = 1 - \frac{3}{2}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}} F(\eta-\beta),$$

or, approximately,
$$\zeta = 1 - \frac{3}{2}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}} \frac{\sqrt{\pi}}{2} e^{\eta-\beta}. \tag{5·19}$$

An expression for $(\eta-\beta)$ in terms of T and θ' may be obtained by using the relation

$$F(\eta+\beta) = \frac{4}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}.$$

For $(\eta+\beta) < 0$, this gives, approximately,

$$\frac{\sqrt{\pi}}{2} e^{\eta+\beta}\left\{1 - \frac{e^{\eta+\beta}}{2^{\frac{3}{2}}}\right\} = \frac{4}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}.$$

Substituting for $(\eta+\beta)$ in (5·19),

$$\zeta = 1 - 2\left\{1 + \frac{2}{3}\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}\right\} e^{-2\theta'/T}, \tag{5·20}$$

which reduces, in the limit, to the classical result

$$\zeta = 1 - 2e^{-2\theta/T} \sim 1 - 2e^{-2\theta'/T}.$$

This expression will be appropriate for $\epsilon_0/k\theta' \to 0$ For $(\eta+\beta) \gg 1$,

$$\frac{2}{3}(\eta+\beta)^{\frac{3}{2}} \sim \frac{4}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}},$$

and substitution in (5·19) gives

$$\zeta = 1 - \frac{3}{2}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}} \frac{\sqrt{\pi}}{2} \exp\left\{-\frac{2\epsilon_0}{kT}\left(\frac{k\theta'}{\epsilon_0} - 2^{-\frac{1}{3}}\right)\right\}. \tag{5·21}$$

The index of the exponential remains negative provided that $k\theta'/\epsilon_0 > 2^{-\frac{1}{3}}$, the condition previously given, (5·10) for $\zeta_0 = 1$. It would not be expected that this expression would pass over into (5·20) for increasing values of $k\theta'/\epsilon_0$, for the two expressions are derived from approximations to $F(\eta+\beta)$ which break down completely in the range between $(\eta+\beta) \gg 1$ and $(\eta+\beta) < 0$. The dominating factor in both expressions, however, is the exponential. Writing (5·21) in the form,

$$\zeta = 1 - \frac{3\sqrt{\pi}}{4}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}} \exp\left[-2\left(\frac{\theta'}{T}\right)\left\{1 - 2^{-\frac{1}{3}}\left(\frac{\epsilon_0}{k\theta'}\right)\right\}\right],$$

it is evident that the rate of decrease of the magnetization with increase of temperature near absolute zero is greater the smaller $k\theta'/\epsilon_0$ At the critical value of $k\theta'/\epsilon_0$ for complete saturation ($\zeta_0 = 1$) at absolute zero, the approximate expression is

$$\zeta = 1 - \frac{3\sqrt{\pi}}{4}\left(\frac{kT}{\epsilon^0}\right)^{\frac{3}{2}} = 1 - \frac{3}{4}\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\left(\frac{T}{\theta'}\right)^{\frac{3}{2}}. \tag{5·22}$$

This expression, it should be emphasized, is of very specialized applicability, and even for the value of $k\theta'/\epsilon_0$ for which it applies ($k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$) the temperature range of approximate validity will be very restricted. It does, however, mark the transition from the range of values of $k\theta'/\epsilon_0$ for which the approach to saturation is exponential in character to that in which it may appropriately be represented by a power series.

For $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$, $\zeta_0 < 1$, under these conditions, for $kT/\epsilon_0 \to 0$, $(\eta+\beta) \gg 1$ and $(\eta-\beta) \gg 1$, and formulae for the variation of magnetization near absolute zero may be obtained from the relation (cf. (5·9) and (5·10))

$$\frac{(1+\zeta_0)^{\frac{2}{3}} - (1-\zeta_0)^{\frac{2}{3}}}{2\zeta_0} = \frac{(1+\zeta)^{\frac{2}{3}} - (1-\zeta)^{\frac{2}{3}}}{2\zeta}$$

$$+ \frac{\pi^2}{12}\cdot\left(\frac{kT}{\epsilon_0}\right)^2\cdot\frac{(1+\zeta)^{-\frac{2}{3}} - (1-\zeta)^{-\frac{2}{3}}}{2\zeta}.$$

Writing $\zeta_0 - \zeta = x$, for $x/\zeta_0 \to 0$, the result obtained may be put in the form

$$\frac{x}{\zeta_0} = \frac{1}{2}\cdot\frac{27}{2}\cdot\frac{\pi^2}{12}\cdot\frac{1}{\zeta_0^2}\cdot\left(\frac{kT}{\epsilon_0}\right)^2 f(\zeta_0), \tag{5·23}$$

where

$$f(\zeta) = \frac{2\zeta}{27}\left\{\frac{(1+\zeta)^{\frac{2}{3}} - (1-\zeta)^{\frac{2}{3}}}{(1-\zeta^2)^{\frac{2}{3}}}\right\} \Big/ \left\{\frac{1}{3(1-\zeta^2)^{\frac{1}{3}}} - \frac{(1+\zeta)^{\frac{2}{3}} - (1-\zeta)^{\frac{2}{3}}}{2\zeta}\right\}.$$

The value of $f(\zeta_0)$ varies little from unity for the range $0 \leqslant \zeta_0 \leqslant 0·9$. For $\zeta_0 \to 0, f(\zeta_0) \to 1$; the value decreases slowly to a minimum of about 0·991 for ζ_0 between 0·6 and 0·7, increases to 1·044 for $\zeta_0 = 0·9$, and subsequently more rapidly (see Table III A). Since $x/\zeta_0 \to 0$, the above result may be written, for $\zeta/\zeta_0 \to 1$,

$$\left(\frac{\zeta}{\zeta_0}\right)^2 = 1 - \frac{27}{2} \cdot \frac{\pi^2}{12} \cdot \frac{1}{\zeta_0^2} \cdot \left(\frac{kT}{\epsilon_0}\right)^2 f(\zeta_0), \tag{5·24}$$

For $\zeta_0 \to 0$, substitution from (5·12) then gives

$$(\zeta/\zeta_0)^2 = 1 - (T/\theta)^2, \tag{5·25}$$

as an expression valid for $\zeta/\zeta_0 \to 1$ as well as for $\zeta/\zeta_0 \to 0$, as shown above. For $k\theta/\epsilon_0 \to 0$, therefore, the variation of magnetization over the whole range $0 \leqslant kT/\epsilon_0 \leqslant k\theta/\epsilon_0$ is represented by the simple expression (5·25). It should be noted that the $k\theta/\epsilon_0$ range for which the expression (5·24) is a valid representation of the approach to saturation is much wider than that for which the simple relation (5·12) between θ and ζ_0 holds. Since an expression in terms of (T/θ) is more immediately illuminating than one involving ζ_0 and kT/ϵ_0, it may finally be noted that for the whole range for which

$$\zeta_0 < 1 \; (k\theta'/\epsilon_0 < 2^{-\frac{1}{2}})$$

the temperature dependence of the magnetization for $\zeta/\zeta_0 \to 1$ may be represented by

$$(\zeta/\zeta_0)^2 = 1 - \alpha(T/\theta)^2, \tag{5·26}$$

where

$$\alpha = \frac{27}{2} \cdot \frac{\pi^2}{12} \cdot \left(\frac{k\theta}{\epsilon_0 \zeta_0}\right)^2 f(\zeta_0),$$

of approximate value 1. (The subsequent computational results show that up to $\zeta_0 = 0·8$, $\alpha < 1·1$) For comparison with the computational results, the following form is useful

$$(T/\theta) = \gamma\{1 - (\zeta/\zeta_0)^2\}^{\frac{1}{2}}. \tag{5·27}$$

From the above treatment, for $\zeta_0 < 1$,

$$\text{for } \zeta/\zeta_0 \to 0, \;\; \gamma \to 1,$$

$$\text{for } \zeta/\zeta_0 \to 1, \;\; \gamma \to \frac{2\sqrt{2}}{3\pi} \cdot \left(\frac{\epsilon_0 \zeta_0}{k\theta}\right) \cdot \frac{1}{\sqrt{f(\zeta_0)}}$$

$$= 0·300\,105 \frac{\zeta_0}{k\theta/\epsilon_0} \cdot \frac{1}{\sqrt{f(\zeta_0)}}.$$

Values of γ obtained from this formula are given in Table III A; the computed values are given in Table III and shown graphically in fig. 3 (see p. 403).

6. Computational details and results

(i) *Relation between* ζ *and* (T/θ) *for* $(\epsilon_0/k\theta) \to 0$

In the limit $\epsilon_0/k\theta \to 0$ the present treatment gives the results of the simple quantum modification of the Weiss molecular field theory (or, equivalently, of the first approximation in the Heisenberg treatment) Table I gives the corresponding values of ζ as a function of T/θ, and of T/θ (and also θ/T) as a function of ζ. An adequate table of this kind seems desirable in itself, and it serves here the particular purpose of providing a standard of comparison for the results as a whole. A four-place table of ζ as a function of T/θ has been given previously (Stoner 1931); the values have been recomputed with higher precision, by the method indicated in § 4 (equation (4·5)), using the tables of Hayashi (1926) for the hyperbolic functions, and the usual methods of interpolation. The values of T/θ as a function of ζ can be readily obtained (see equation (4·6)) as Hayashi has tabulated the function $\tanh^{-1} x$ Either method of tabulation, by itself, has drawbacks as a representation of the functional relation between ζ and T/θ, but the two methods together enable the whole range to be covered fairly adequately without extremely small intervals. The awkward ranges, in either form of the table, may be dealt with more conveniently by the use of series expansions, than by tabulation at very small intervals.

TABLE I. MAGNETIZATION AS A FUNCTION OF TEMPERATURE, AND TEMPERATURE AS A FUNCTION OF MAGNETIZATION, IN THE LIMIT $(\epsilon_0/k\theta) \to 0$ CORRESPONDING TO CLASSICAL STATISTICS

ζ, relative magnetization, θ, Curie temperature.

T/θ	ζ	ζ	T/θ	θ/T
0·0	1·000 000	0·0	1·000 000	1·000 000
0·1	1·000 000	0·1	0·996 658	1·003 353
0·2	0·999 909	0·2	0·986 521	1·013 663
0·3	0·997 414	0·3	0·969 244	1·031 732
0·4	0·985 624	0·4	0·944 178	1·059 122
0·50	0·957 504	0·50	0·910 239	1·098 612
0·55	0·935 529	0·55	0·889 419	1·124 330
0·60	0·907 337	0·60	0·865 617	1·155 245
0·65	0·872 065	0·65	0·838 387	1·192 767
0·70	0·828 635	0·70	0·807 102	1·239 001
0·75	0·775 516	0·75	0·770 848	1·297 273
0·80	0·710 412	0·80	0·728 191	1·373 265
0·85	0·629 501	0·85	0·676 669	1·477 827
0·90	0·525 429	0·90	0·611 322	1·635 799
0·95	0·379 484	0·95	0·518 621	1·928 190
1·00	0·000 000	1·00	0·000 000	∞

(ii) $k\theta'/\epsilon_0$ *as a function of* kT/ϵ_0 *and* ζ

In the method followed here the primary computations are those of the reduced interaction energy coefficient, $k\theta'/\epsilon_0$, as a function of the reduced temperature, kT/ϵ_0, and the relative magnetization, ζ. The fundamental procedure is that described in § 4 It may be illustrated by details of the computation of $k\theta'/\epsilon_0$ for $kT/\epsilon_0 = 0{\cdot}4$, $\zeta = 0{\cdot}0$ (A) and $0{\cdot}5$ (B).

A. $\underline{kT/\epsilon_0 = 0{\cdot}4,\ \zeta = 0{\cdot}0.}$

$$F(\eta+\beta) = F(\eta-\beta) = F(\eta) = \frac{2}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}} = 2{\cdot}63523.$$

By inverse interpolation in the $F(\eta)$ table (McDougall and Stoner 1938),

$$\eta = 2{\cdot}10087.$$

By direct interpolation in the $F'(\eta)$ table,

$$F'(\eta) = 1{\cdot}33483.$$

Whence
$$\theta'/T = F/F' = 1{\cdot}97420,$$
$$k\theta'/\epsilon_0 = 0{\cdot}78968.$$

It may be noted that this value of $k\theta'/\epsilon_0$ corresponds to a Curie temperature θ, where $k\theta/\epsilon_0 = 0{\cdot}4$.

B. $\underline{kT/\epsilon_0 = 0{\cdot}4,\ \zeta = 0{\cdot}5.}$

$$F(\eta+\beta) = \frac{2}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}(1+\zeta) = 3{\cdot}95285,$$

$$F(\eta-\beta) = \frac{2}{3}\left(\frac{\epsilon_0}{kT}\right)^{\frac{3}{2}}(1-\zeta) = 1{\cdot}31762.$$

By inverse interpolation,

$$\eta+\beta = 2{\cdot}98528, \quad \eta-\beta = 0{\cdot}91178.$$

Whence,
$$\beta = 1{\cdot}03675,$$
$$\theta'/T = \beta/\zeta = 2{\cdot}07350,$$
$$k\theta'/\epsilon_0 = 0{\cdot}82940.$$

The values of $k\theta'/\epsilon_0$ for $\zeta = 0{\cdot}0$ were obtained by the procedure A at intervals of $0{\cdot}1$ over the range $0{\cdot}1 \leqslant kT/\epsilon_0 \leqslant 2{\cdot}0$. The procedure B was used for $\zeta = 0{\cdot}1, 0{\cdot}2, \ldots, 0{\cdot}9, 0{\cdot}95$ for the range $0{\cdot}1 \leqslant kT/\epsilon_0 \leqslant 1{\cdot}0$. The method used for inverse interpolation was the two-machine method of Comrie (1936). In the course of this and subsequent stages of the work many hundreds of inverse interpolations must be made. The method described in connexion with the tables of Fermi-Dirac functions (McDougall and Stoner 1938), involving the use of an inverse Taylor series, is appropriate for immediate application to the values tabulated ($F(\eta)$ and derivatives up to $F'''(\eta)$). The method has advantages over those involving direct interpolation for estimated interpolates, requires only one machine, and is not unduly time consuming if only a few values are needed. If many inverse interpolations are required, however, the Comrie method is very much quicker, even if it is necessary to draw up difference tables for the tabulated function. The method is particularly convenient if differences beyond the third may be neglected, or if the fourth difference is sufficiently small that it may be taken into account by modifying the second difference by the "throw-back" method. In most of the present work the first and third and the modified second differences were used, and usually the error in the inverse interpolate was then within that due to uncertainty in the last digit of the tabulated function. Under these conditions the time required for inverse interpolation, using two machines, is little greater than that for direct interpolation. Use is made of the values of the Besselian interpolation coefficients tabulated by Comrie. For full details of the method reference may be made to Comrie's account.

For the range in which the required inverse interpolates were most concentrated, it was found advantageous to draw up an inverse table, giving η at equal intervals of $F(\eta)$, and to use direct interpolation for the particular values required. This table covered the range $0{\cdot}30 \leqslant F(\eta) \leqslant 6{\cdot}0$ (corresponding to $1{\cdot}03593 \leqslant \eta \leqslant 4{\cdot}11739$) with the following ranges and intervals: $0{\cdot}30-1{\cdot}00$ ($0{\cdot}02$); $1{\cdot}00-2{\cdot}50$ ($0{\cdot}05$); $2{\cdot}5-6{\cdot}0$ ($0{\cdot}1$). Outside this range, inverse interpolation was carried out directly using the η, $F(\eta)$ table.

The values of $k\theta'/\epsilon_0$ for $kT/\epsilon_0 = 0$ for the different values of ζ (these particular values are referred to as ζ_0) were found by the use of equation (5·10) and are given in the first line of Table II.

For $kT/\epsilon_0 \geqslant 1{\cdot}1$, the values of $k\theta'/\epsilon_0$ are more readily calculated by the method indicated in connexion with equations (5·6) and (5·7). The value for $\zeta = 0{\cdot}5$, for example, is obtained by adding to the computed value of (θ'/T) for $\zeta = 0{\cdot}0$ the "classical" difference between (θ'/T) for $\zeta = 0{\cdot}0$ and $\zeta = 0{\cdot}5$ (obtained from Table I), and the small correction term in $(\epsilon_0/kT)^{\frac{3}{2}}\zeta^2$ of equation (5·7). A comparison with the values obtained by the standard method for $kT/\epsilon_0 = 1{\cdot}0$ shows that the maximum error in θ'/T in the range covered will not exceed 0·5 in the fifth decimal place (this being the error for

$kT/\epsilon_0 = 1{\cdot}0$, $\zeta = 0{\cdot}95$) and will usually be very much less. The values of $k\theta'/\epsilon_0$ for $\zeta = 0{\cdot}0$ for $kT/\epsilon_0 \geqslant 1{\cdot}0$ were computed both by the standard method and by the series (5·6). At $kT/\epsilon_0 = 1{\cdot}0$, the two values differed by 0·5 in the fifth place. As kT/ϵ_0 increases, the precision attainable by the standard method decreases (it is limited by the accuracy to which $F'(\eta)$ is known) while that of the series method (using a fixed number of terms—those of equation (5·6)) increases. The series values are slightly high, the error will be approximately proportional to $(\epsilon_0/kT)^6$, and assuming that it is 0·5 in the fifth place at $kT/\epsilon_0 = 1{\cdot}0$, it will become less than 1 in the sixth place for $kT/\epsilon_0 > 1{\cdot}4$. By the use of the two methods the values of $k\theta'/\epsilon_0$ over the range $1{\cdot}0 \leqslant kT/\epsilon_0 \leqslant 2{\cdot}0$ were obtained with a probable accuracy of within 2 or 3 units in the sixth place, the rounded five-place values given should be correct to one unit in the last place.

The interval of 0·1 in kT/ϵ_0 can be reduced to 0·05 without difficulty for $kT/\epsilon_0 > 0{\cdot}4$. For smaller values of kT/ϵ_0, however, the higher differences are large, and interpolation is inaccurate. Values of $k\theta'/\epsilon_0$ were, therefore, also computed for $kT/\epsilon_0 = 0{\cdot}05$, 0·15, 0·25 and 0·35, the last agreeing throughout to within 0·3 in the fifth place with the values found by interpolation. For $kT/\epsilon_0 = 0{\cdot}05$ the values of $F(\eta + \beta)$, except for $\zeta = 0{\cdot}0$, were outside the range of the $F(\eta)$ table, and the values of $(\eta + \beta)$ were obtained by the use of the series relation (2·16). A check on the values for $\zeta = 0{\cdot}0$ is obtained by the use of the series relation (5·3) for $k\theta'/\epsilon_0$. (This formula (5·3), it may be noted, though convenient, is of limited applicability It is satisfactory up to $kT/\epsilon_0 = 0{\cdot}05$, the error then being within 1 in the sixth place, but even for $kT/\epsilon_0 = 0{\cdot}10$, the error is about 4 in the fifth place)

The rounded values given in Table II are believed to be correct to 1 unit in the last (fifth) place. This is probably the limit which can be consistently attained throughout the table, in so far as the values are derived from the published $F(\eta)$ values. Using the values on which these were based the accuracy is greater—probably to within 2 units in the sixth place; a check by the method of differences indicates that this degree of accuracy is attained in the values on which Table II is based, but little useful purpose would be served by encumbering Table II with an additional doubtful digit in each entry. It may be added that the computation of values of $k\theta'/\epsilon_0$ for $\zeta > 0{\cdot}95$ would in general be somewhat laborious, as $F(\eta + \beta)$ and $F(\eta - \beta)$ would usually lie outside the range of the tables, and two rather troublesome series calculations would be required for each value of $k\theta'/\epsilon_0$.

The general character of the relations shown numerically in the table is shown graphically in fig. 1, in which $k\theta'/\epsilon_0$ is plotted against kT/ϵ_0 for $\zeta = 0{\cdot}0$, 0·5, 0·8 and 0·95. The corresponding classical relations are also shown by the straight lines of slope θ/T (see Table I). This figure shows in perhaps the most immediately significant way the character of the modification resulting from the replacement of classical by Fermi-Dirac statistics. The curve for $\zeta = 0{\cdot}0$ is of particular interest as it gives the relation between the reduced Curie temperature ($k\theta/\epsilon_0 = kT/\epsilon_0$ for $\zeta = 0{\cdot}0$) and the magnitude of the interaction energy coefficient. (It corresponds to the first column of Table II.) Moreover, the height of this curve ($k\theta'/\epsilon_0$) is proportional to the inverse of the susceptibility in the absence of interchange interaction; and by lowering the curve by an amount $k\theta'/\epsilon_0$, so that it intersects the kT/ϵ_0 axis at $k\theta/\epsilon_0$, the curve to the right of the intersection gives a representation of the variation with temperature of the inverse of the susceptibility above the Curie point.

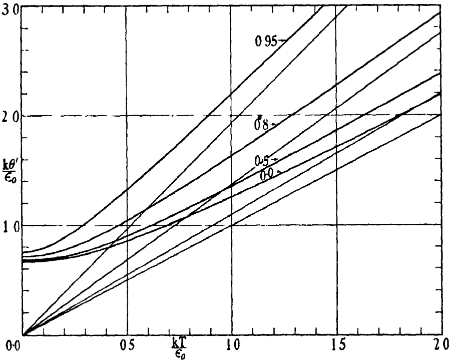


FIG. 1. $k\theta'/\epsilon_0$ as a function of kT/ϵ_0 and ζ for $\zeta = 0{\cdot}0$, 0 5, 0·8 and 0 95. $k\theta'$, interaction energy coefficient. ϵ_0, maximum Fermi zero point energy, without interaction. ζ, relative magnetization. The numbers on the curves give the values of ζ. The straight lines are asymptotes to the curves and represent the classical relations between θ and T which hold for $\epsilon_0/k\theta' \to 0$.

For $k\theta'/\epsilon_0 < 2/3$, spontaneous magnetization does not occur, while for $2/3 < k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$, $0 < \zeta_0 < 1$ where ζ_0 is the maximum value of the relative magnetization, that at absolute zero. This effect has no classical analogue.

TABLE II. $k\theta'/\epsilon_0$ AS A FUNCTION OF kT/ϵ_0 AND ζ

$k\theta'$, interaction energy coefficient; ϵ_0, maximum Fermi zero point energy, without interaction; ζ, relative magnetization. $\zeta = M/N\mu$, where M is the magnetic moment, N the total number of electrons, μ the Bohr magneton. (For $\epsilon_0/k\theta' \to 0$, $\theta \to \theta'$, where θ is the Curie temperature.)

	ζ										
kT/ϵ_0	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	0·95
0·00	0·66667	0·66716	0·66867	0·67127	0·67511	0·68041	0·68758	0·69733	0·71108	0·73255	0·75006
0·05	0·66805	0·66856	0·67010	0·67276	0·67669	0·68213	0·68953	0·69964	0·71412	0·73765	0·75891
0·10	0·67244	0·67298	0·67464	0·67751	0·68177	0·68774	0·69597	0·70753	0·72485	0·75583	0·78779
0·15	0·68067	0·68129	0·68319	0·68650	0·69146	0·69849	0·70837	0·72260	0·74472	0·78642	0·83157
0·20	0·69369	0·69443	0·69669	0·70065	0·70660	0·71512	0·72722	0·74488	0·77281	0·82655	0·88571
0·25	0·71166	0·71254	0·71524	0·71996	0·72709	0·73732	0·75193	0·77340	0·80759	0·87393	0·94741
0·30	0·73410	0·73513	0·73830	0·74384	0·75223	0·76431	0·78157	0·80703	0·84769	0·92688	1·01484
0·35	0·76032	0·76151	0·76517	0·77157	0·78127	0·79524	0·81524	0·84477	0·89201	0·98417	1·08666
0·40	0·78968	0·79103	0·79519	0·80247	0·81351	0·82940	0·85217	0·88582	0·93969	1·04489	1·16195
0·45	0·82162	0·82314	0·82780	0·83597	0·84835	0·86619	0·89176	0·92955	0·99008	1·10835	1·24000
0·5	0·85569	0·85737	0·86254	0·87161	0·88534	0·90513	0·93351	0·97546	1·04268	1·17404	1·32029
0·6	0·92885	0·93087	0·93707	0·94792	0·96438	0·98810	1·02212	1·07241	1·15301	1·31059	1·48605
0·7	1·00708	1·00943	1·01665	1·02931	1·04850	1·07616	1·11582	1·17447	1·26849	1·45229	1·65698
0·8	1·08900	1·09168	1·09993	1·11439	1·13631	1·16792	1·21324	1·28026	1·38769	1·59774	1·83166
0·9	1·17370	1·17672	1·18600	1·20226	1·22692	1·26247	1·31345	1·38884	1·50969	1·74598	2·00914
1·0	1·26054	1·26390	1·27421	1·29228	1·31968	1·35917	1·41581	1·49957	1·63385	1·89639	2·18878
1·1	1·34909	1·35278	1·36412	1·38400	1·41413	1·45757	1·51987	1·61201	1·75971	2·04850	2·37013
1·2	1·43900	1·44302	1·45539	1·47708	1·50995	1·55734	1·62530	1·72581	1·88694	2·20198	2·55285
1·3	1·53002	1·53438	1·54779	1·57128	1·60689	1·65823	1·73185	1·84074	2·01528	2·35658	2·73669
1·4	1·62198	1·62667	1·64111	1·66640	1·70475	1·76004	1·83933	1·95659	2·14456	2·51211	2·92146
1·5	1·71471	1·71974	1·73521	1·76231	1·80340	1·86263	1·94758	2·07322	2·27462	2·66842	3·10701
1·6	1·80811	1·81347	1·82997	1·85888	1·90271	1·96589	2·05651	2·19052	2·40534	2·82540	3·29322
1·7	1·90207	1·90778	1·92530	1·95602	2·00258	2·06972	2·16599	2·30838	2·53663	2·98294	3·48001
1·8	1·99653	2·00257	2·02112	2·05365	2·10295	2·17403	2·27598	2·42674	2·66841	3·14098	3·66728
1·9	2·09142	2·09779	2·11738	2·15171	2·20375	2·27878	2·38638	2·54552	2·80062	3·29944	3·85498
2·0	2·18668	2·19339	2·21400	2·25014	2·30492	2·38390	2·49717	2·66468	2·93321	3·45828	4·04306

It is shown in fig. 1 by the different heights at which the different ζ curves meet the $k\theta'/\epsilon_0$ axis. The relation between $k\theta'/\epsilon_0$ and ζ_0 is shown in fig. 2. The curve corresponds to the first line of Table II, but is plotted so as to bring out the character of the inverse relation.

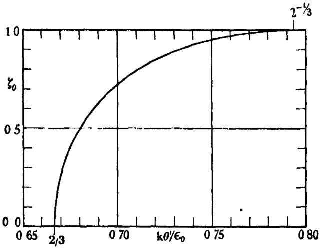


FIG. 2. ζ_0 as a function of $k\theta'/\epsilon_0$. ζ_0, relative magnetization at absolute zero. For $\zeta_0 = 0\ 0$, $k\theta'/\epsilon_0 = 2/3$, for $\zeta_0 = 1\cdot0$, $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$.

(iii) *kT/ϵ_0 as a function of $k\theta'/\epsilon_0$ and ζ*

The variation with temperature of the spontaneous magnetization of a material characterized by a particular value of the interaction energy coefficient $k\theta'$ and the Fermi zero point energy ϵ_0 may be determined by inverse interpolation of Table II. It is necessary to determine, for each value of ζ, the value of kT/ϵ_0 corresponding to the particular value of $k\theta'/\epsilon_0$. (This corresponds, graphically, to a determination of the values of kT/ϵ_0 at the points of intersection of horizontal constant $k\theta'/\epsilon_0$ line with the curves of fig. 1.) The precision with which the inverse interpolation may be made is limited, ultimately, by the precision with which the $k\theta'/\epsilon_0$ values have been computed. For the higher values of kT/ϵ_0 for each value of ζ in Table II this attainable precision is reached by the method of inverse interpolation described above. For the lower values, however, the higher differences become so large that the method is inaccurate. This difficulty affects mainly the range of $k\theta'/\epsilon_0$ for which the relative magnetization at absolute zero, ζ_0, is less than unity, and it is convenient to consider this range separately, partly for this reason, and partly because a method of presenting the results slightly different from that used for the higher values of $k\theta'/\epsilon_0$ seems appropriate.

(*a*) $2/3 \leqslant k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{3}}$.

As has been shown, spontaneous magnetization does not occur for $k\theta'/\epsilon_0 < \frac{2}{3}$, while for $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$, ζ_0 is less than unity. For this range it is appropriate to consider the magnetization relations for suitably spaced values of ζ_0 rather than of $k\theta'/\epsilon_0$; the values of kT/ϵ_0 have therefore been determined for different values of ζ (the values in Table II) for a series of values of ζ_0 (0·0, 0·1, 0·2, ...). The results are given in Table III. For the value of $k\theta'/\epsilon_0$ for which ζ_0 reaches the value unity ($k\theta'/\epsilon_0 = 2^{-\frac{1}{3}} = 0{\cdot}793701$) the value of kT/ϵ_0 for which $\zeta = 0$ is 0·4065. As this is the highest kT/ϵ_0 value for the range, inverse interpolation is required entirely in the first part of Table II. In this region the first differences are small (as will be immediately apparent from the curves of fig. 1) which limits the attainable precision of the inverse interpolate, and the higher differences are large (or not available for the first few entries) which renders the simple method of inverse interpolation inaccurate. The second difficulty could be overcome by reducing the intervals in Table II, but as this cannot be done accurately by interpolation in this part of the table, a very considerable number of additional computations (of the type A and B of (ii)) would be required. The procedure adopted was to make inverse interpolations by the simple method (extrapolating the differences where necessary on the assumption, of approximate validity, that the functions are even in the argument kT/ϵ_0 near $kT/\epsilon_0 = 0$) and to recalculate $k\theta'/\epsilon_0$ from the approximate values of kT/ϵ_0 so obtained. The values of $d(k\theta'/\epsilon_0)/d(kT/\epsilon_0)$ can be found with sufficient accuracy from Table II (using Comrie's tables for computing derivatives from differences) to correct the approximate values of (kT/ϵ_0). The approximate and corrected values occasionally differed by 1 in the fourth decimal place for $kT/\epsilon_0 < 0{\cdot}1$, but usually by much less.

The values of $k\theta'/\epsilon_0$ corresponding to given values of ζ_0 can be found (by equation (5·10)) without difficulty (using Barlow's tables) to seven decimal places; but in general the values of $k\theta'/\epsilon_0$ on which those given in Table II are based cannot claim an accuracy greater than 1 in the sixth place. This applies in particular to the values for $\zeta = 0$, the accuracy of which depends on that of $F'(\eta)$. (The accuracy of the values for $\zeta > 0$ depends on that of $F(\eta+\beta) - F(\eta-\beta)$, which is in most cases considerably greater than $F'(\eta)$, but has the same absolute uncertainty, so that the relative uncertainty is considerably smaller.) Assuming an uncertainty of 1 in the sixth place in the $k\theta'/\epsilon_0$ values, there is an unavoidable uncertainty in the kT/ϵ_0 values found by inverse interpolation, which would range, for $\zeta = 0$, from about 3 in the fifth place for the first entry in the table ($kT/\epsilon_0 = 0{\cdot}0300$) to about 2 in the sixth place for the last ($kT/\epsilon_0 = 0{\cdot}4065$). With reference to the first of these, it seems somewhat surprising, though the reason is obvious on consideration, that a value can be given correctly to only 3 significant figures when it is derived from functions (the $F(\eta)$) which have been evaluated correctly to 8 significant figures (using 9 or 10 in the computations). The values for kT/ϵ_0 on which those in Table III are based for $\zeta_0 \geqslant 0{\cdot}3$ are, however, probably correct to at least 1 in the fifth place, and for the few entries for $\zeta_0 < 0{\cdot}3$, more accurate evaluations of the basic functions have been made, using the series (2·13) and (2·15). Not only, therefore, should the values of kT/ϵ_0 be correct to within 1 in the fourth place (as given), but also the other values tabulated (which are based on the six place values of kT/ϵ_0) should not be in error by more than 1 in the last place. It should be stated that owing to the peculiar character of the variation of the functions the application of the method of differences to the tabulated values in this range can in general give only a very rough check.

The second entry in the columns $0 < \zeta < \zeta_0$ for each value of ζ_0 gives T/θ, or $(kT/\epsilon_0)/(k\theta/\epsilon_0)$, $k\theta/\epsilon_0$ being the value of kT/ϵ_0 for $\zeta = 0$. The third entry gives $\gamma = (T/\theta)/\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$ (see equation 5·27), and so gives an immediate indication of the extent by which the form of the magnetization-temperature curve differs from that given by the simple formula,

$$(T/\theta)^2 = 1-(\zeta/\zeta_0)^2. \tag{6·1}$$

It has been shown that for the range in which $\zeta_0 < 1$, $\gamma \to 1$ for $(\zeta/\zeta_0) \to 0$, and further that γ approaches 1 for the whole range of values of ζ/ζ_0 as ζ_0 approaches zero. This is borne out by the computed values in the table. A theoretical formula for γ for $\zeta \to \zeta_0$ has also been given (equation (5·27)); the corresponding values (using the computed values of $k\theta/\epsilon_0$) are shown in Table III A, these serving to supplement the computed values for $\zeta < \zeta_0$ shown in Table III. The set of values is sufficient to indicate the character of the deviations from the simple magnetization-temperature relation (6·1) over the whole range. Some of the results are shown graphically in fig. 3. (Since γ has been determined only for integral values of 10ζ the form of the curves is not determined with any precision in the ζ/ζ_0 range between $(\zeta_0 - 0{\cdot}1)/\zeta_0$ and 1 for the higher values of ζ_0. The curves do not provide information additional to that in the tables, but serve to bring out its character.) Curves are not shown for the smaller values of ζ_0 (0·1, 0·2 and 0·3) but it is evident from the values in the tables that these lie below that for $\zeta_0 = 0{\cdot}4$. For all values of ζ/ζ_0, γ at first increases and then decreases with ζ_0, but the change occurs in such a way that for $\zeta_0 \geqslant 0{\cdot}6$ the variation over the ζ/ζ_0 range is not monotonic, and a minimum appears in the curves.

TABLE III. kT/ϵ_0 AS A FUNCTION OF ζ_0 AND ζ FOR THE RANGE $\frac{2}{3} < k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{3}} = 0{\cdot}793701$

ζ, relative magnetization; ζ_0, relative magnetization at absolute zero. The three entries for each value of ζ_0 in the columns $0{\cdot}0 < \zeta < \zeta_0$ give (1) kT/ϵ_0, (2) T/θ, (3) $(T/\theta)/\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$. In the column $\zeta = 0{\cdot}0$, the value of kT/ϵ_0 is given, this corresponding to $k\theta/\epsilon_0$, where θ is the Curie temperature. The values corresponding to (2) and (3) would be 1·0000 throughout and are not entered. For $\zeta = \zeta_0$, the values (not entered) corresponding to (1) and (2) would be 0·0000 throughout. The ratio (3) takes the indeterminate form 0/0; the limiting values to which the ratio tends as $\zeta \to \zeta_0$ are discussed in the text, and given in Table III A.

		ζ										
ζ_0	$k\theta'/\epsilon_0$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	0·95
0·1	0·667162	0·0300										
0·2	0·668673	0·0600	0·0520									
			0·8664									
			1·0004									
0·3	0·671275	0·0899	0·0848	0·0671								
			0·9433	0·7467								
			1·0005	1·0018								
0·4	0·675108	0·1194	0·1156	0·1036	0·0793							
			0·9687	0·8677	0·6646							
			1·0005	1·0020	1·0048							
0·5	0·680410	0·1488	0·1458	0·1365	0·1193	0·0898						
			0·9799	0·9172	0·8018	0·6038						
			1·0001	1·0007	1·0022	1·0063						
0·6	0·687581	0·1789	0·1764	0·1685	0·1546	0·1330	0·0990					
			0·9857	0·9417	0·8641	0·7435	0·5532					
			0·9997	0·9988	0·9978	0·9975	1·0008					
0·7	0·697330	0·2113	0·2090	0·2020	0·1898	0·1716	0·1457	0·1070				
			0·9891	0·9558	0·8982	0·8123	0·6894	0·5066				
			0·9993	0·9974	0·9942	0·9898	0·9850	0·9835				
0·8	0·711083	0·2486	0·2464	0·2398	0·2286	0·2121	0·1893	0·1584	0·1143			
			0·9913	0·9649	0·9196	0·8532	0·7617	0·6372	0·4600			
			0·9992	0·9966	0·9920	0·9852	0·9757	0·9633	0·9502			
0·9	0·732552	0·2968	0·2947	0·2883	0·2775	0·2616	0·2402	0·2118	0·1742	0·1222		
			0·9929	0·9714	0·9348	0·8815	0·8091	0·7136	0·5870	0·4116		
			0·9991	0·9963	0·9915	0·9840	0·9731	0·9573	0·9339	0·8984		
0·95	0·750059	0·3312	0·3291	0·3227	0·3118	0·2960	0·2746	0·2466	0·2099	0·1606	0·0876	
			0·9936	0·9743	0·9414	0·8937	0·8291	0·7444	0·6337	0·4850	0·2646	
			0·9992	0·9966	0·9922	0·9853	0·9751	0·9602	0·9373	0·8993	0·8265	
1·0	0·793701	0·4065	0·4043	0·3976	0·3863	0·3698	0·3476	0·3186	0·2810	0·2310	0·1599	0·1077
			0·9946	0·9782	0·9502	0·9097	0·8552	0·7839	0·6912	0·5683	0·3934	0·2649
			0·9996	0·9983	0·9961	0·9926	0·9875	0·9798	0·9679	0·9472	0·9025	0·8484

TABLE III A. SOME MAGNITUDES CHARACTERIZING THE MAGNETIZATION RELATIONS FOR THE RANGE $2/3 \leqslant k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{2}}$

For explanation of the symbols ζ, ζ_0, and θ, and for the values of $k\theta'/\epsilon_0$ and $k\theta/\epsilon_0$, see Table III. The second column shows the deviation from linearity in the relation between $k\theta/\epsilon_0$ and ζ_0. ($2\sqrt{2}/3\pi = 0{\cdot}300105$.) The values in the column $(\zeta/\zeta_0)^2/\{1-(T/\theta)\}$ are approximations to the negative slope of the $(\zeta/\zeta_0)^2$, (T/θ) curve at the Curie point. The coefficient γ is defined by

$$T/\theta = \gamma\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}.$$

In the calculation of the values for $\zeta/\zeta_0 \to 1$, $f(\zeta_0)$ is required, and also the values in column 2. (See equations (5·23) to (5·27).)

ζ_0	$\dfrac{3\pi}{2\sqrt{2}}\cdot\dfrac{(k\theta/\epsilon_0)}{\zeta_0}$	$\dfrac{(\zeta/\zeta_0)^2}{1-(T/\theta)}$ for $\zeta/\zeta_0 = 0{\cdot}1$	$f(\zeta_0)$	γ for $\zeta/\zeta_0 \to 1$
0·0	(1·0000)	(2·00)	(1·0000)	(1·0000)
0·1	0·9999	2·00	0·9993	1·0004
0·2	0·9995	2·02	0·9986	1·0012
0·3	0·9980	2·03	0·9969	1·0036
0·4	0·9945	2·04	0·9947	1·0082
0·5	0·9915	2·02	0·9928	1·0122
0·6	0·9936	1·95	0·9912	1·0109
0·7	1·0059	1·88	0·9925	0·9979
0·8	1·0353	1·81	1·0022	0·9648
0·9	1·0989	1·74	1·0442	0·8905
0·95	1·1617	1·74	1·1230	0·8123
1·0	1·3545	1·84	—	—

TABLE III B. APPROXIMATE VALUES OF T/θ AS A FUNCTION OF ζ/ζ_0 FOR $k\theta/\epsilon_0 \to 0$. IN THIS LIMIT $T/\theta \sim \{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$

ζ/ζ_0	$\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$
0·0	1·000 000
0·1	0·994 987
0·2	0·979 796
0·3	0·953 939
0·4	0·916 515
0·50	0·866 025
0·55	0·835 165
0·60	0·800 000
0·65	0·759 934
0·70	0·714 143
0·75	0·661 438
0·80	0·600 000
0·85	0·526 783
0·90	0·435 890
0·95	0·312 250
1·00	0·000 000

The indications are that this minimum persists, becoming more pronounced, and occurring closer to $\zeta/\zeta_0 = 1$ as ζ_0 increases in the range of $k\theta'/\epsilon_0$ corresponding to $\zeta_0 < 1$. The approximate representation of the magnetization curves by a formula of the type (6·1) becomes increasingly less appropriate. The result to be mainly emphasized, however, is that even up to $\zeta_0 = 0{\cdot}7$, γ never differs from unity by as much as 2 %, so that in this range, the very simple expression (6·1), considered as a formula for (ζ/ζ_0) gives a surprisingly close approximation to the magnetization as a function of (T/θ) over the whole temperature range in which spontaneous magnetization occurs. In view of this approximate validity of the formula (6·1) over a fairly extended range, the values of $\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$ are given in Table III A, this table being complementary to Table I. It may be noticed that in the limit here considered, (T/θ) is the same function of (ζ/ζ_0) as is (ζ/ζ_0) of (T/θ), so that one set of pairs of values corresponds to the two sets in Table I.

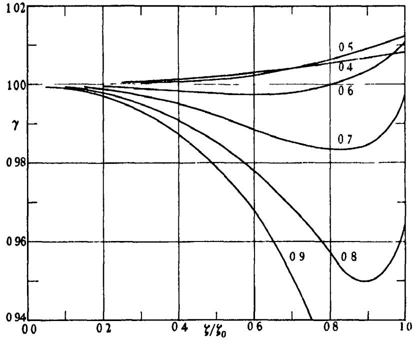


FIG. 3. Curves showing the deviation of the magnetization-temperature relation from that given by the formula $(\zeta/\zeta_0)^2 = 1-(T/\theta)^2$. $\gamma = (T/\theta)/\{1-(\zeta/\zeta_0)^2\}^{\frac{1}{2}}$; ζ, relative magnetization; ζ_0, relative magnetization at absolute zero. The numbers on the curve give the values of ζ_0.

The second column in Table III A shows that the relation between ζ_0 and $k\theta/\epsilon_0$ is linear to within 1 % up to $\zeta_0 = 0{\cdot}7$, there being agreement to this extent with the approximate formula (5·13). At first the ratio $(k\theta/\epsilon_0\zeta_0)$

decreases, as indicated by (5·12), though rather more rapidly (illustrating the failings of formulae based on the asymptotic series expansions), and subsequently increases. The values in the third column give an approximation to the slope of the curve at the Curie point. (These were determined by calculating the kT/ϵ_0 values corresponding to $\zeta/\zeta_0 = 0{\cdot}1$, first approximations being obtained by interpolation from Table III. For reasons already discussed, high precision cannot be claimed for these values.) The slope at first increases slightly from the limiting value 2, decreases to a minimum, and finally increases; this increase continues with increase of $k\theta'/\epsilon_0$ beyond the range under immediate consideration, eventually approaching the value 3 in the classical limit (corresponding to $\epsilon_0/k\theta' \to 0$)

(*b*) $k\theta'/\epsilon_0 > 2^{-\frac{1}{3}} = 0{\cdot}793701$.

In the range of values of $k\theta'/\epsilon_0$ in which the interchange interaction is sufficient to produce complete saturation at absolute zero ($\zeta_0 = 1$), it is appropriate to consider the character of the magnetization curves for a series of suitably spaced values of $k\theta'/\epsilon_0$. In Table IV are given the values of kT/ϵ_0 for $\zeta = 0{\cdot}0, 0{\cdot}1, \ldots, 0{\cdot}9$ and $0{\cdot}95$ for values of $k\theta'/\epsilon_0$ at intervals of 0·1 from 0·8 to 2·0. These are obtained by inverse interpolation from Table II. The differences are such that, except for the smaller values of $k\theta'/\epsilon_0$ and the larger values of ζ, the values of kT/ϵ_0 obtained by the use of the Comrie two-machine method as described above should be as accurate as the precision of the tabulated values of Table II allows. This was tested by carrying out a direct calculation (by the procedure B above) with the kT/ϵ_0 values found for $\zeta = 0{\cdot}9$ and $0{\cdot}95$ for the $k\theta'/\epsilon_0$ range 0·8 to 1·2. It was found that the corrected and approximate values agreed to within 1 unit in the sixth place, except for $k\theta'/\epsilon_0 = 0{\cdot}8$, showing that in general in the range covered by Table IV it is unnecessary to carry out the lengthy "recalculations" which are required in connexion with Table III. The values in Table IV for kT/ϵ_0 should be correct to 1 unit in the fifth place (as given), and the other values tabulated (based on six-place kT/ϵ_0 values) should not be in error by more than 1 or 2 units in the last place. The method of differences, however, is applicable only as a rough check, particularly for the entries corresponding to the higher ζ and lower $k\theta'/\epsilon_0$ values.

The second entry in the columns $\zeta > 0{\cdot}0$ for each value of $k\theta'/\epsilon_0$ gives T/θ. The third entry gives the ratio of T/θ to the value $(T/\theta)_{cl}$ corresponding to the classical limit (Table I). The sequence of values of these ratios for a given $k\theta'/\epsilon_0$ gives an immediate indication of the character of the magnetization temperature curve by comparison with the standard classical curve, while the manner of approach to the classical type of curve is shown by the sequence for given values of ζ. The values of T/θ for a given ζ are smaller than the classical values, and the ratio $(T/\theta)/(T/\theta)_{cl}$ is smaller the larger the value of ζ. For a given ζ, the ratio decreases more and more rapidly with decreasing $k\theta'/\epsilon_0$, becoming zero at the values of $k\theta'/\epsilon_0$ shown in the first line of Table II. The variation for a given value of ζ/ζ_0 is the same as that for a given ζ down to the limiting value for which $\zeta_0 = 1$ ($k\theta'/\epsilon_0 = 0{\cdot}793701$), below this value of $k\theta'/\epsilon_0$, however, the ratio falls to a minimum and subsequently increases, and, after passing through a slight maximum, finally attains a limiting value corresponding to the T/θ value shown in Table III B. (The occurrence of the secondary maximum as well as the minimum is indicated by the numbers on which the curves of fig 3 are based.)

In detail the variation in the character of the relation between T/θ and ζ/ζ_0 as $k\theta'/\epsilon_0$ decreases from high values to the limiting value of $\frac{2}{3}$ (corresponding to a decrease of $k\theta/\epsilon_0$ from high values to zero) is remarkably complex The salient points are that the ζ/ζ_0, T/θ curves always fall below the classical curve, the difference being most pronounced (if the ratio of the T/θ values is considered) for the larger values of ζ, the change proceeds monotonically until $k\theta'/\epsilon_0$ reaches the critical value which is just sufficient to produce saturation at absolute zero; the curves for lower values of $k\theta'/\epsilon_0$ may lie in part or wholly above the curve for this critical $k\theta'/\epsilon_0$, finally approaching the form indicated by the numbers in Table III B.

(iv) *Graphical representations*

The general character of the relations expressed by the values in Tables III and IV is most readily appreciated with the aid of graphical illustrations. In fig. 4 a number of curves are given showing ζ as a function of kT/ϵ_0 for a series of values of $k\theta'/\epsilon_0$. The falling curves give the spontaneous magnetization below the Curie point, while the rising curves on the right give $(\mu H/\zeta\epsilon_0)$, a quantity proportional to the inverse of the susceptibility above the Curie point. As these quantities are obtained simply by subtracting $k\theta'/\epsilon_0$ from the values given in the column $\zeta = 0{\cdot}0$ in Table II (see equation (3·8)) it is unnecessary to tabulate them explicitly. These inverse susceptibility curves approach asymptotically a series of parallel lines, of slope unity, passing through the points $kT/\epsilon_0 = k\theta'/\epsilon_0$ on the kT/ϵ_0 axis.

Since with the simple form of the classical statistical treatment ζ/ζ_0 is a unique function of T/θ, it has become usual to plot these quantities as found experimentally with ferromagnetics with a view to comparison with theory. With Fermi-Dirac statistics, the curves differ according to the value of $k\theta'/\epsilon_0$; a series of such curves is shown in fig. 5. The values of $k\theta'/\epsilon_0$ correspond to those in fig. 4, except that the classical curve ($k\theta'/\epsilon_0 \to \infty$) is drawn in place of

TABLE IV. kT/ϵ_0 AS A FUNCTION OF $k\theta'/\epsilon_0$ AND ζ FOR THE RANGE $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}} = 0{\cdot}793701$

ζ, relative magnetization. In this range the relative magnetization at absolute zero, ζ_0, has the value unity. The three entries for each value of $k\theta'/\epsilon_0$ in the columns $\zeta > 0{\cdot}0$ give (1) kT/ϵ_0, (2) T/θ, (3) $(T/\theta)/(T/\theta)_{cl}$. The classical values $(T/\theta)_{cl}$ are those corresponding to $\epsilon_0/k\theta' \to 0$, and are given in Table I. In the column $\zeta = 0{\cdot}0$, the value of kT/ϵ_0 is given, this corresponding to $k\theta/\epsilon_0$, where θ is the Curie temperature. The values corresponding to (2) and (3) would be 1·00000 throughout, and are not entered.

	ζ										
$k\theta'/\epsilon_0$	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	0·95
0·8	0·41657	0·41435	0·40761	0·39616	0·37959	0·35723	0·32802	0·29008	0·23977	0·16813	0·11546
		0·99467	0·97850	0·95101	0·91122	0·85757	0·78744	0·69636	0·57560	0·40360	0·27718
		0·99801	0·99187	0·98119	0·96510	0·94213	0·90969	0·86279	0·79045	0·66021	0·53445
0·9	0·56151	0·55897	0·55126	0·53814	0·51915	0·49355	0·46006	0·41652	0·35861	0·27519	0·21208
		0·99547	0·98174	0·95838	0·92457	0·87896	0·81933	0·74179	0·63865	0·49010	0·37769
		0·99881	0·99516	0·98879	0·97923	0·96564	0·94653	0·91908	0·87704	0·80170	0·72826
1·0	0·69116	0·68827	0·67950	0·66458	0·64298	0·61381	0·57565	0·52592	0·45958	0·36328	0·28930
		0·99582	0·98313	0·96154	0·93028	0·88809	0·83287	0·76093	0·66493	0·52561	0·41857
		0·99915	0·99656	0·99205	0·98528	0·97567	0·96217	0·94279	0·91313	0·85979	0·80709
1·1	0·81316	0·80991	0·80008	0·78333	0·75908	0·72633	0·68342	0·62746	0·55263	0·44352	0·35901
		0·99601	0·98392	0·96332	0·93350	0·89322	0·84046	0·77164	0·67961	0·54543	0·44150
		0·99935	0·99736	0·99389	0·98869	0·98130	0·97093	0·95606	0·93329	0·89222	0·85130
1·2	0·93052	0·92693	0·91602	0·89746	0·87056	0·83422	0·78660	0·72442	0·64115	0·51937	0·42457
		0·99614	0·98442	0·96447	0·93556	0·89651	0·84533	0·77851	0·68902	0·55815	0·45627
		0·99948	0·99787	0·99507	0·99088	0·98492	0·97656	0·96457	0·94621	0·91302	0·87978
1·3	1·04477	1·04083	1·02886	1·00848	0·97895	0·93905	0·88672	0·81835	0·72671	0·59239	0·48748
		0·99622	0·98477	0·96526	0·93700	0·89880	0·84872	0·78328	0·69556	0·56701	0·46659
		0·99956	0·99822	0·99589	0·99240	0·98744	0·98048	0·97049	0·95519	0·92751	0·89967

1·4	1·15679	1·15250	1·13947	1·11729	1·08514	1·04168	0·98467	0·91016	0·81019	0·66345	0·54855
		0·99629	0·98503	0·96585	0·93806	0·90049	0·85121	0·78680	0·70037	0·57352	0·47420
		0·99963	0·99849	0·99650	0·99352	0·98929	0·98335	0·97484	0·96180	0·93817	0·91434
1·5	1·26714	1·26250	1·24842	1·22444	1·18968	1·14268	1·08101	1·00038	0·89213	0·73305	0·60826
		0·99634	0·98522	0·96630	0·93887	0·90178	0·85311	0·78948	0·70405	0·57851	0·48003
		0·99968	0·99869	0·99696	0·99438	0·99071	0·98555	0·97817	0·96685	0·94632	0·92558
1·6	1·37618	1·37119	1·35606	1·33029	1·29293	1·24241	1·17610	1·08938	0·97289	0·80154	0·66694
		0·99637	0·98538	0·96665	0·93950	0·90279	0·85461	0·79160	0·70695	0·58244	0·48463
		0·99972	0·99884	0·99733	0·99505	0·99182	0·98729	0·98079	0·97083	0·95275	0·93446
1·7	1·48419	1·47885	1·46267	1·43512	1·39517	1·34113	1·27020	1·17741	1·05271	0·86915	0·72480
		0·99641	0·98551	0·96694	0·94002	0·90361	0·85583	0·79330	0·70929	0·58561	0·48835
		0·99975	0·99897	0·99762	0·99560	0·99272	0·98869	0·98290	0·97404	0·95794	0·94163
1·8	1·59134	1·58566	1·56844	1·53910	1·49657	1·43903	1·36350	1·26465	1·13178	0·93606	0·78201
		0·99643	0·98561	0·96717	0·94044	0·90429	0·85682	0·79471	0·71121	0·58822	0·49141
		0·99977	0·99908	0·99786	0·99604	0·99346	0·98984	0·98464	0·97668	0·96221	0·94754
1·9	1·69780	1·69178	1·67351	1·64240	1·59728	1·53626	1·45613	1·35125	1·21022	1·00239	0·83867
		0·99645	0·98570	0·96737	0·94080	0·90485	0·85766	0·79588	0·71281	0·59040	0·49398
		0·99979	0·99916	0·99807	0·99642	0·99408	0·99080	0·98610	0·97888	0·96578	0·95248
2·0	1·80366	1·79730	1·77800	1·74511	1·69742	1·63291	1·54819	1·43729	1·28813	1·06822	0·89488
		0·99647	0·98577	0·96753	0·94110	0·90533	0·85836	0·79687	0·71418	0·59225	0·49615
		0·99981	0·99924	0·99824	0·99674	0·99461	0·99162	0·98733	0·98075	0·96880	0·95666

that for $k\theta'/\epsilon_0 = 1{\cdot}4$. For $T/\theta < 1$, all the curves corresponding to values of $k\theta'/\epsilon_0$ greater than that required to produce saturation lie between those shown for $k\theta'/\epsilon_0 = 0{\cdot}794$ and ∞; and similarly for the curves corresponding to the inverse of the susceptibility. The inverse susceptibility curves approach parallel asymptotes, of slope unity, passing through the points θ'/θ on the T/θ axis. An adequate representation of the differences in the ζ/ζ_0 curves for values of $k\theta'/\epsilon_0$ such that $\zeta_0 < 1$ is hardly possible in this form; the curve shown for $\zeta_0 = 0{\cdot}5$ differs inappreciably for $T/\theta < 1$ from the limiting curve for $\zeta_0 \to 0$. Although the curves below the Curie point for this range of $k\theta'/\epsilon_0$ values are very similar, the inverse susceptibility curves are sharply distinguished from those for the range for which $\zeta_0 = 1$, and from each other, in that the slope of the asymptotes is equal to ζ_0 instead of to unity; this is due to the absolute saturation moment (corresponding to complete alignment of the available electron spins) being greater than the attainable saturation moment at absolute zero.

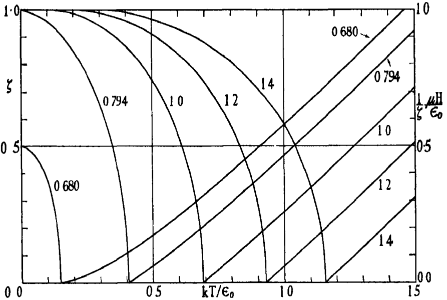


FIG. 4. Spontaneous magnetization below and inverse of susceptibility above the Curie point as functions of kT/ϵ_0. ζ, relative magnetization. The numbers on the curves give the values of $k\theta'/\epsilon_0$. The two curves on the left correspond to $k\theta'/\epsilon_0$ values giving $\zeta_0 = 0{\cdot}5$ ($k\theta'/\epsilon_0 = 0{\cdot}680410$) and the critical value for complete saturation at absolute zero, $\zeta_0 = 1{\cdot}0$ ($k\theta'/\epsilon_0 = 0{\cdot}793701$). The falling curves (scale on left) give the spontaneous magnetization below the Curie point, the rising curves (scale on right) a quantity proportional to the inverse of the susceptibility ($\chi \propto \zeta/H$).

The curves of the type shown in fig. 5 (or the numbers on which they are based) are particularly appropriate for purposes of comparison with experi-

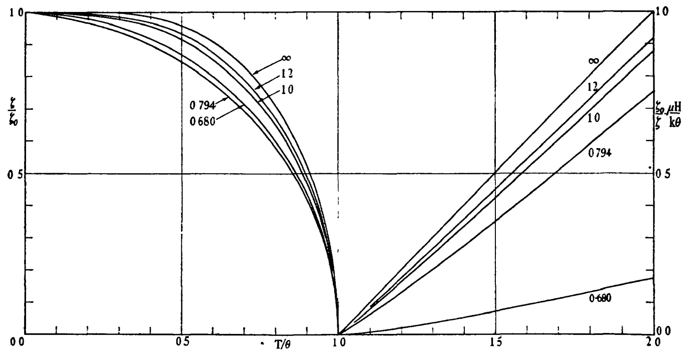


FIG. 5. Curves showing the reduced magnetization below and the reduced inverse susceptibility above the Curie point as a function of T/θ. ζ, relative magnetization; ζ_0, relative magnetization at absolute zero. The numbers on the curves give the values of $k\theta'/\epsilon_0$. (See also caption of fig. 4.) The limiting curves labelled ∞ correspond to the use of classical statistics.

ment, for the data required for the construction of the corresponding experimental curves are obtainable directly from measurements of magnetic characteristics (i.e. without a knowledge of the Fermi zero point energy, ϵ_0, which would be required in constructing experimental curves of the type shown in fig. 4). The theoretical curves give an immediate indication of the nature of the differences in observable magnetic characteristics which may be associated with differences in the ratio of the interaction energy coefficient, $k\theta'$, to the Fermi energy, ϵ_0.

7. GENERAL DISCUSSION

The results of the present treatment of ferromagnetism, as embodied in Tables I–IV of § 6, and in the formulae of §§ 3–5, are, formally, sets of relations consistent with the premises involved in the initial equations (§ 2). A comparison of the results with those found experimentally should provide evidence as to whether the premises are appropriate as approximate representations of the physical conditions in ferromagnetics. The value of the treatment is not, however, to be measured simply from the degree of agreement obtainable in a casual comparison. The value lies rather in the completeness of the scheme obtained of possible magnetic characteristics consistent with basic relations of a very simple type, these relations being of a kind indicated by more general theory, the development of which is independent, at least in principle, of the particular applications being made.

Any adequate discussion of experimental results in relation to the scheme would involve the consideration of other factors, such as those concerned in the various "domain" effects in ferromagnetics, which would rather confuse the immediate questions at issue besides extending the paper very considerably. A straightforward comparison of the theoretical results for the variation of magnetization with temperature below the Curie point cannot be made, for the experiments do not yield directly, with any certainty, the values of the spontaneous magnetization, with which the theory deals (cf. Stoner 1936*c*). Even if allowance could easily be made for domain effects, however, it would be desirable to defer the fuller consideration of the experimental results until a more rigorous treatment of electronic specific heats in ferromagnetics has been developed. This would enable further information to be obtained about the Fermi zero point energy, ϵ_0, which, as far as the magnetic results alone are concerned, has rather the character of a parameter adjustable to obtain the best fit. Although use has been made of provisional estimates of ϵ_0 for nickel derived from the specific heat (cf. Stoner 1936*d*) a much more satisfactory treatment of this problem can be developed without undue difficulty on the basis of the computational results which have now been obtained. With this additional information it may be anticipated that a much less arbitrary comparison with experiment will be possible.

Excluding detailed consideration, a few general comments on the relation between the theoretical and experimental results are appropriate. In previous comparisons of the experimental magnetization curves below the Curie point of nickel, iron and cobalt (Tyler 1930, 1931, Stoner 1931) with the classical theoretical curve (the curve labelled ∞ in fig. 5), it has been found that there is general agreement in form. The experimental curve, however, falls more rapidly at low temperatures and near the Curie point, crossing over the theoretical curve in the region of T/θ values of 0·6–0·7. The low temperature discrepancy is brought out by the fact that the experimental results are better represented by a law of approach involving T^2 or $T^{3/2}$ rather than $\exp(-\theta/T)$. This is in line with the present treatment. It is only for very large values of the interchange interaction (or strictly of the ratio $k\theta'/\epsilon_0$) that the Fermi-Dirac curve approximates at low temperatures to that given by classical statistics, while for lower values of $k\theta'/\epsilon_0$ the approach to saturation is better represented by a law involving a power of T, ultimately T^2. Near the Curie point the curves for different values of $k\theta'/\epsilon_0$ do not differ greatly from each other; but since all the curves lie below the classical curve for the whole range, while the experimental curve lies above for the last part of the range, it is apparent that other factors must be involved than those considered in this treatment. It is possible that the discrepancy arises from neglect of domain effects, a neglect analogous in some respects to neglect of local order effects in order-disorder phenomena in alloys, the consideration of which leads to important modifications in the Curie point region.

For temperatures above the Curie point the difficulties connected with distinguishing between the "true" and "apparent" magnetization disappear. The curvature of the $1/\chi$, T lines which is unmistakably shown by the observations, and which is inexplicable on the basis of classical statistics, is an immediate consequence of the present treatment; at the same time, the approximate linearity over long ranges, which is also established experimentally, emerges from the detailed computations (see figs. 4, 5). Moreover, it follows from the theory that the value of the saturation moment deduced from the paramagnetism above the Curie point (by applying the usual Weiss law relations) will in general differ from that indicated by the magnetization at low temperatures. This will hold even when there is complete saturation at absolute zero ($\zeta_0 = 1$), for the slopes of the approximately linear portions of the $1/\chi$, T curves are in general less than those of the asymptotes (so that higher saturation moments would be inferred), and the apparent discrepancy will be even greater if complete saturation is not attained ($\zeta_0 < 1$).

Attention has previously been drawn to the insufficiency of a positive interaction effect as a condition for ferromagnetism (Stoner 1936*a*); the interaction energy coefficient must exceed a certain critical value ($k\theta'/\epsilon_0 > \frac{2}{3}$). The further point that saturation at absolute zero is incomplete unless a second critical value is exceeded ($k\theta'/\epsilon_0 > 2^{-\frac{1}{3}}$) is of particular interest, as showing that the number of "holes" in the electron bands concerned may be appreciably greater than that indicated by the magnetization. This possibility has been noticed and briefly discussed by Slater (1936*b*). The quantitative treatment of this particular effect given here, it may be noted, follows entirely from an elementary application of Fermi-Dirac statistics, though the precise forms of the magnetization temperature curves can be determined only by the rather elaborate computations.

It is perhaps desirable to refer again explicitly to the fact that the particularities in many of the relations obtained, such as that between $k\theta'/\epsilon_0$ and ζ_0 (see fig. 2) are dependent on the form of energy band assumed, in which the energy density of states is proportional to the square root of the energy, measured from an appropriate zero. It seems unlikely, however, that the general character of the results should be critically dependent on this precise form. The treatment has the advantages that it enables the changes to be followed over the complete range of parameters which would be relevant in almost any treatment, and that it links up with the familiar and useful Weiss treatment of ferromagnetism (which is, in fact, a special case), with the theory of spin paramagnetism of free electrons, and with the general theory of metals as treated by the method of collective electron energy bands.

I have pleasure in acknowledging my indebtedness to the Government Grants Committee of the Royal Society for the loan of a Brunsviga calculating machine.

SUMMARY

After a brief survey of the problem of ferromagnetism (§ 1), general equations are obtained, using Fermi-Dirac statistics, for the magnetic moment M of a number N of electrons in an unfilled energy band of standard form, for which the interchange interaction effects give rise to a term in the

energy expression proportional to the square of the magnetization (§ 2). The relative magnetization, ζ ($= M/N\mu$), is expressible as an implicit function of the reduced field, $\mu H/\epsilon_0$, temperature, kT/ϵ_0 and interaction energy coefficient, $k\theta'/\epsilon_0$, where ϵ_0 is the Fermi zero energy without interchange interaction. Particular limits correspond to previously considered equations for electron spin paramagnetism, and the Weiss-Heisenberg "classical" equations for ferro- and paramagnetism.

The treatment of paramagnetism above the Curie point, θ, is straightforward (§ 3). The general method of dealing with the spontaneous magnetization below the Curie point (§ 4) is to obtain first values of $k\theta'/\epsilon_0$ corresponding to a series of values of kT/ϵ_0 and ζ (Table II, fig. 1). From these, values of kT/ϵ_0 as a function of ζ for a series of values of $k\theta'/\epsilon_0$ are obtained by inverse interpolation (Tables III, IV, figs. 4, 5) (§ 6). The computational work is based largely on the use of tabulated values of Fermi-Dirac functions (McDougall and Stoner 1938).

The character of the dependence of ζ on kT/ϵ_0 or on T/θ depends on the ratio $k\theta'/\epsilon_0$, the classical results being obtained for $\epsilon_0/k\theta' \rightarrow 0$. A necessary condition for ferro-magnetism is $k\theta'/\epsilon_0 > \frac{2}{3}$, while for $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}} = 0{\cdot}793701$, the relative magnetization at absolute zero, ζ_0, is less than unity (fig. 2). For small values of ζ_0 the magnetization-temperature curve is closely represented by

$$(\zeta/\zeta_0)^2 = 1 - (T/\theta)^2,$$

but the curve does not change monotonically to the classical form (Table I) as $k\theta'/\epsilon_0$ increases (fig. 5) It is not possible to describe briefly the rather complex changes in detail of the curves (§§ 5, 6).

In addition to the numerical results, series expansions appropriate for particular ranges are given for $k\theta'/\epsilon_0$ as a function of $k\theta/\epsilon_0$, and of kT/ϵ_0 and ζ; and expressions are derived for the variation of magnetization near the Curie point, and at low temperatures (§ 5).

A full discussion of the experimental results is deferred, but it is shown that a major difficulty is removed in that the lack of agreement in the values deduced for the saturation moment from the paramagnetism above the Curie point and from the low temperature magnetization receives an immediate interpretation on the basis of the theoretical scheme (§ 7).

REFERENCES

Comrie, L. J. 1936 "Interpolation and Allied Tables." Reprinted from the *Nautical Almanac* for 1937: H.M. Stationery Office.

Hayashi, K. 1926 "Sieben- und mehrstellige Tafeln der Kreis- und Hyperbelfunktionen." Berlin: Springer.

McDougall, J. and Stoner, E. C. 1938 *Philos. Trans.* A, **237**, 67.
Mott, N. F. 1935 *Proc. Phys. Soc.* **47**, 571.
Mott, N. F. and Jones, H. 1936 "The Theory of the Properties of Metals and Alloys." Oxford.
Slater, J. C. 1936 *a* *Phys. Rev.* **49**, 537.
— 1936 *b* *Phys. Rev.* **49**, 931.
Stoner, E. C. 1931 *Phil. Mag.* **12**, 737.
— 1933 *Phil. Mag.* **15**, 1018.
— 1935 *Proc. Roy. Soc.* A, **152**, 672.
— 1936 *a* *Proc. Roy. Soc.* A, **154**, 656.
— 1936 *b* *Proc. Leeds Phil. Soc.* **3**, 191.
— 1936 *c* *Philos. Trans.* A, **235**, 165.
— 1936 *d* *Phil. Mag.* **22**, 81.
Tyler, F. 1930 *Phil. Mag.* **9**, 1026.
— 1931 *Phil. Mag.* **11**, 596.

Resonance in crystal beams of sodium-ammonium seignette salt

By W. MANDELL, PH.D.
Chelsea Polytechnic, London

(*Communicated by O. W. Richardson, F.R.S.—Received* 13 *November* 1937)

INTRODUCTION

It has been noticed by several observers that, with longitudinal oscillations in certain crystal beams, on keeping the length constant and gradually decreasing either the breadth or thickness, a sudden change of considerable magnitude may occur in the wave-length. This has been attributed to some resonance effect. Lack (1929) describes interesting results obtained with rectangular quartz beams of 30° cut. The X axis, as with Voigt, coincided with an electric axis and constituted the length direction of the beam. The breadth was parallel to the optic axis. With such a beam there is a low vibration frequency depending upon the length, and a high vibration frequency varying inversely as the thickness, the field being applied in the direction of the thickness. A square plate was used, and on gradually reducing the thickness, several discontinuities occurred on plotting the wave-length against the thickness. According to Lack these discontinuities occurred at frequencies that could be identified with harmonics of the frequency which the crystal would have if it were vibrating in the direction of the length. That is, the longitudinal oscillation along the length could affect the frequency which was supposed to depend only on the thickness, so that the system acted as two coupled electrical circuits, although the two sets of oscillations were in planes at right angles.

Somewhat similar results have been obtained for quartz beams by Hitchcock (1930). He used the Curie (or perpendicular) cut, where the length is perpendicular to an electric axis and the thickness parallel to it, whilst the breadth is parallel to the optic axis Hitchcock showed that the frequency of oscillations along the length suddenly changed in value on decreasing the breadth.

It will be shown that the beams used in this paper are suitable for the study of this phenomenon, owing to their unique elastic and piezo-electric properties.

THE PIEZO-ELECTRIC PROPERTIES OF THE CRYSTALS OF SEIGNETTE SALT

The crystals used were of sodium-ammonium seignette salt

$$(NaNH_4C_4H_4O_6 4H_2O).$$

The theory of piezo-electricity applicable to all crystals has been completely solved by Voigt (1910), and a short account of the modification of this general theory applicable to these crystals has been mentioned by the writer in a former paper (Mandell 1928*a*). There are three piezo-electric moduli only, namely

$$d_{14} = +56{\cdot}0 \times 10^{-8},\ d_{25} = -149{\cdot}5 \times 10^{-8},\ \text{and}\ d_{36} = +28{\cdot}3 \times 10^{-8}.$$

Fig. 1 shows the orientation of the beams cut from the crystal and used in the present experiments. X, Y and Z represent the a, b and c crystallographic axes and are all at right angles. Three planes are now drawn each containing a pair of axes so that the planes also are mutually perpendicular. Three rectangular beams may now be cut as shown. Consider any one of the planes shown, say $XOYL$. The dotted line OL bisects the angle XOY. In the case of the beam shown, (*de*) parallel to OL represents its length, (*fe*) the breadth, and (*fh*) the thickness. This will be the convention used in the present paper, namely, for all of the three planes the length of the beam will always be along the direction of the bisector of the angle between the two axes considered. The breadth will always lie in the plane and will be at right angles to the length, while the thickness will be measured in the direction at right angles to the plane, that is, it will always be perpendicular to the plane and parallel to a crystallographic axis.

Since there are only three piezo-electric moduli, the general equations of Voigt (1910, p. 902) reduce to

$$y_z = d_{14}E_1, \quad z_x = d_{25}E_2, \quad \text{and} \quad x_y = d_{36}E_3.$$

The third equation, for instance, states that if an electric field E_3 be applied along the direction of the Z axis, then in the perpendicular plane XY, a shear strain is produced. If therefore the length (OL) of the beam bisects the angle XOY, longitudinal displacements take place along the length and along the breadth.

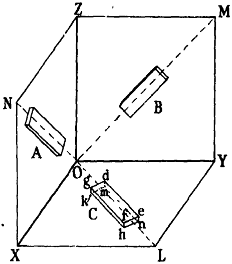


Fig. 1

On referring again to fig 1, if light metal plates be placed on the upper and lower surfaces (*gdef*) and (*kmnh*) and a steady potential difference be applied to the plates, the crystal will either extend along (*de*) and contract along (*fe*), or contract along (*de*) and extend along (*fe*), according to the sign of the electric field. If an alternating field be now substituted then the crystal may oscillate longitudinally in one of two directions, either along the length, or along the breadth, on condition that the frequency of the applied field is exactly equal to either of these natural frequencies which are determined only by the elastic constants and dimensions of the beam. All three beams act in a similar manner and if electric fields of the required frequency be applied in the direction of a crystallographic axis, that is, perpendicular to a plane, longitudinal oscillations will occur either along the length or along the breadth.

THE LATERAL CONTRACTIONS ACCOMPANYING AN INCREASE IN LENGTH FOR BEAMS FROM CRYSTALS OF THE RHOMBIC SYSTEM

At a later stage reference will be made to the elastic properties of the beams so that their values are determined here. The geometrical nature of elastic moduli is clearly set out by Voigt who shows that the elastic properties of a crystal are completely characterized by the combination of a bitensor and a tensor surface. For crystals of the most complex type, namely, those belonging to the triclinic system, there are twenty-one elastic moduli. The bitensor system is built up by combining the components of four vectors and supplies fifteen terms, each term being a usual component of a bitensor surface while the tensor system supplies a further six terms. The expression for the thermodynamic potential is then transformed into an algebraic expression of twenty-one terms representing those two surfaces, the coefficient of each term being one of the elastic moduli or a combination of two of them, and at the same time a component of the two surfaces. The bitensor surface is therefore a fourth power expression in x, y and z and the tensor surface, a second power expression. With the aid of these two surfaces any elastic modulus for a beam cut in any direction can easily be determined. Suppose that a new set of rectangular axes X', Y', Z', are related to the original axes X, Y, Z, by the scheme of direction cosines shown:

$$\begin{array}{cccc} & X' & Y' & Z' \\ X & \alpha_1 & \beta_1 & \gamma_1 \\ Y & \alpha_2 & \beta_2 & \gamma_2 \\ Z & \alpha_3 & \beta_3 & \gamma_3 \end{array}$$

Then Voigt (1910, p. 590) shows that

$$\begin{aligned} s_{11}^{1} = {} & s_{11}\alpha_1^4 + s_{22}\alpha_2^4 + s_{33}\alpha_3^4 + (2s_{23}+s_{44})\alpha_2^2\alpha_3^2 + (2s_{31}+s_{55})\alpha_3^2\alpha_1^2 + (2s_{12}+s_{66})\alpha_1^2\alpha_2^2 \\ & + 2(s_{14}+s_{56})\alpha_1^2\alpha_2\alpha_3 + 2(s_{25}+s_{64})\alpha_2^2\alpha_3\alpha_1 + 2(s_{36}+s_{45})\alpha_3^2\alpha_1\alpha_2 \\ & + 2\alpha_1^3(s_{15}\alpha_3 + s_{16}\alpha_2) + 2\alpha_2^3(s_{26}\alpha_1 + s_{24}\alpha_3) + 2\alpha_3^3(s_{34}\alpha_2 + s_{35}\alpha_1). \end{aligned} \tag{1}$$

This expression gives the extension modulus along the length when tensile stresses are applied in this direction, which coincides with the new X' axis. This complex expression simplifies considerably when it is remembered that due to the symmetry of the crystal considered, twelve of the elastic moduli vanish, whilst the remaining nine have already been determined by a statical method (Mandell 1928*b*, p. 122). Thus for beam C, the expression reduces to

$$\begin{aligned} s_{11}^1 &= s_{11}\alpha_1^4+s_{22}\alpha_2^4+(2s_{12}+s_{66})\,\alpha_1^2\alpha_2^2 \\ &= 1/4(s_{11}+s_{22}+2s_{12}+s_{66}) = 47{\cdot}8\times 10^{-10}, \end{aligned} \tag{2}$$

since the angles measured in degrees are $\alpha_1 = 45$, $\alpha_2 = 45$, and $\alpha_3 = 90$.

Suppose now, in beam C of fig. 1, the X' axis now along (OL), be rotated through a further 90 degrees in the plane $(OXLY)$, so that it is parallel to the breadth, then on substituting the appropriate values for the direction cosines in this equation we obtain the same numerical value for the extension for an equal tensile stress, one gram weight per square centimetre, along the direction of the breadth.

In like manner for beam B,

$$s_{22}^1 = s_{22}\beta_2^4+s_{33}\beta_3^4+(2s_{23}+s_{44})\,\beta_2^2\beta_3^2 = 37{\cdot}5\times 10^{-10}, \tag{3}$$

gives the extension modulus for stresses either along the length or the breadth, and for beam A,

$$s_{33}^1 = s_{33}\gamma_3^4+s_{11}\gamma_1^4+(2s_{31}+s_{55})\,\gamma_1^2\gamma_3^2 = 94{\cdot}2\times 10^{-10}, \tag{4}$$

for the two directions.

Again Voigt (1910, p. 592) shows that the general expression for the lateral contraction for a crystal beam referred to the new set of rectangular axes is given by the expression

$$\begin{aligned} s_{23}^1 = {} & s_{11}\beta_1^2\gamma_1^2+s_{22}\beta_2^2\gamma_2^2+s_{33}\beta_3^2\gamma_3^2+s_{23}(\beta_2^2\gamma_3^2+\beta_3^2\gamma_2^2)+s_{31}(\beta_3^2\gamma_1^2+\beta_1^2\gamma_3^2) \\ & +s_{12}(\beta_1^2\gamma_2^2+\beta_2^2\gamma_1^2)+s_{44}\beta_2\gamma_2\beta_3\gamma_3+s_{55}\beta_3\gamma_3\beta_1\gamma_1+s_{66}\beta_1\gamma_1\beta_2\gamma_2 \\ & +s_{56}\beta_1\gamma_1(\beta_2\gamma_3+\beta_3\gamma_2)+s_{64}\beta_2\gamma_2(\beta_3\gamma_1+\beta_1\gamma_3)+s_{45}\beta_3\gamma_3(\beta_1\gamma_2+\beta_2\gamma_1) \\ & +s_{14}(\beta_1^2\gamma_2\gamma_3+\gamma_1^2\beta_2\beta_3)+s_{25}(\beta_2^2\gamma_1\gamma_3+\gamma_2^2\beta_1\beta_3)+s_{36}(\beta_3^2\gamma_2\gamma_1+\gamma_3^2\beta_1\beta_2) \\ & +\beta_1\gamma_1[s_{15}(\beta_1\gamma_3+\gamma_1\beta_3)+s_{16}(\beta_1\gamma_2+\gamma_1\beta_2)] \\ & +\beta_2\gamma_2[s_{26}(\beta_2\gamma_1+\gamma_2\beta_1)+s_{24}(\beta_2\gamma_3+\gamma_2\beta_3)] \\ & +\beta_3\gamma_3[s_{34}(\beta_3\gamma_2+\gamma_3\beta_2)+s_{35}(\beta_3\gamma_1+\gamma_3\beta_1)]. \end{aligned} \tag{5}$$

For crystals of the rhombic system twelve of the elastic moduli again vanish. Let us apply this equation to beam B, where the Y' axis is along the direction of length (OM), the Z' axis being at right angles to this and in the same plane, and the X' and X axes coinciding. Then the angles measured in degrees have the values $\alpha_1 = 0$, $\alpha_2 = 90$, $\alpha_3 = 90$, $\beta_1 = 90$, $\beta_2 = 45$, $\beta_3 = 45$, $\gamma_1 = 90$, $\gamma_2 = 135$, $\gamma_3 = 45$, and

$$\begin{aligned} s_{23}^1 &= s_{22}\beta_2^2\gamma_2^2+s_{33}\beta_3^2\gamma_3^2+s_{23}(\beta_2^2\gamma_3^2+\beta_3^2\gamma_2^2)+s_{44}(\beta_2\gamma_2\beta_3\gamma_3) \\ &= -5{\cdot}3\times 10^{-10}. \end{aligned} \tag{6}$$

This gives the lateral contraction along the direction of breadth for beam B considered as a unit cube, for unit tensile stress along the direction of length. In exactly the same way as before, by considering the rotation of the Y' axis through a further 90° so that it coincides with the direction of breadth it can be shown that for unit tensile stress along the breadth the lateral contraction along the length is $5{\cdot}3 \times 10^{-10}$ for the unit cube.

For the contraction in the direction of the thickness the equation (5) is again used except that the terms are changed in a cyclical manner. Thus the equations reduce to

$$s_{21}^{1} = s_{12}\alpha_1^2\beta_2^2 + s_{31}\alpha_1^2\beta_3^2 = -21{\cdot}05 \times 10^{-10}, \tag{7}$$

$$s_{31}^{1} = s_{31}\gamma_3^2\alpha_1^2 + s_{12}\gamma_2^2\alpha_1^2 = -21{\cdot}05 \times 10^{-10}. \tag{8}$$

The first equation states that for beam B, unit tensile stress along the length for a unit cube gives a contraction along the thickness direction of $21{\cdot}05 \times 10^{-10}$. The second equation states that for a unit tensile stress along the breadth the contraction along the thickness is the same, namely $21{\cdot}05 \times 10^{-10}$.

In a similar way for beam A,

$$\begin{aligned} s_{31}^{1} &= s_{11}\gamma_1^2\alpha_1^2 + s_{33}\gamma_3^2\alpha_3^2 + s_{31}(\gamma_3^2\alpha_1^2 + \gamma_1^2\alpha_3^2) + s_{55}\gamma_3\alpha_3\gamma_1\alpha_1 \\ &= -82{\cdot}2 \times 10^{-10} \end{aligned} \tag{9}$$

is the contraction along the breadth for the extension in length,

$$s_{23}^{1} = s_{23}\beta_2^2\gamma_3^2 + s_{12}\beta_2^2\gamma_1^2 = -6{\cdot}7 \times 10^{-10} \tag{10}$$

the contraction in thickness for extension in length and

$$s_{12}^{1} = s_{12}\alpha_1^2\beta_2^2 + s_{23}\alpha_3^2\beta_2^2 = -6{\cdot}7 \times 10^{-10} \tag{11}$$

is the contraction in thickness for extension in breadth.

Corresponding equations for beam C are

$$s_{12}^{1} = s_{11}\alpha_1^2\beta_1^2 + s_{22}\alpha_2^2\beta_2^2 + s_{12}(\alpha_1^2\beta_2^2 + \alpha_2^2\beta_1^2) + s_{66}\alpha_1\beta_1\alpha_2\beta_2 = -10{\cdot}2 \times 10^{-10} \tag{12}$$

$$s_{31}^{1} = s_{31}\gamma_3^2\alpha_1^2 + s_{23}\gamma_3^2\alpha_2^2 = -19{\cdot}25 \times 10^{-10}, \tag{13}$$

$$s_{23}^{1} = s_{23}\beta_2^2\gamma_3^2 + s_{31}\beta_1^2\gamma_3^2 = -19{\cdot}25 \times 10^{-10}. \tag{14}$$

These equations show, therefore, that all three beams have similar elastic properties.

THE ELECTRICAL CIRCUIT

The electrical circuit used in the experiments was an improved form of that described by Pierce (1923). The crystal was placed across the plate and grid, but an oscillatory circuit replaced the plate resistance. It was found that the grid-leak was not essential and oscillations were much more vigorous without it. On varying the capacity of the condenser in the circuit the crystal started to oscillate strongly and continued to do so with constant frequency through a wide range of the condenser setting. A point of vital importance connected with this arrangement was that the frequency was solely determined by the crystal for there was no measurable change in wave-length on the substitution of other inductances and capacity in the tuned circuit.

The alternating field was applied to the crystal beam by means of two light vertical condenser plates which were made to press lightly against the faces of the beam by means of two fine metal springs. The dimensions of the plates were somewhat larger than the crystal faces to ensure a uniform field. The nature of all oscillations was checked by means of dust-figures with lycopodium powder.

Since the frequency of a crystal is dependent on the temperature, it was necessary to enclose the beam in a thermostat, which was regulated by an electrical device, when measuring some of the small changes in wave-length on changing the lateral dimensions.

THE THEORY OF ELECTRICALLY COUPLED CIRCUITS APPLICABLE TO THE SPECIAL CRYSTAL BEAMS UNDER TEST

Suppose for simplicity that square beams are used, and that the equivalent electrically coupled circuits are as shown in the diagram. The

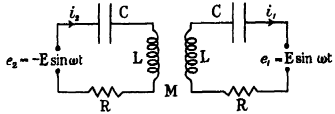


FIG. 2

alternating potentials, assumed of sine form, impressed on the two circuits will differ in phase by π radians, for if the applied field causes an extension in one direction, an equal contraction occurs in the direction at right angles.

On employing the two circuit equations of standard form it may be shown that

$$A = E\sqrt{\frac{\left(\frac{1}{\omega C}-\omega L+\omega M\right)^2+R^2}{\left[\omega^2(M^2-L^2)+\left(\frac{2L}{C}+R^2\right)-\frac{1}{\omega^2 C^2}\right]^2+4R^2\left[\omega L-\frac{1}{\omega C}\right]^2}},$$

gives the amplitude of the current in either circuit.

In order to find the optimum values of ω, for maximum current amplitude, it is convenient to square the expression for the amplitude, differentiate it with respect to ω, and equate to zero. In crystal oscillations the effects of resistance on the period can quite safely be neglected so that the expression obtained then reduces to:

$$[\omega^6C^3(L+M)(L-M)^2+\omega^4C^2(3M+L)(M-L)+\omega^2C(3M-L)+1]$$
$$[\omega^4C^2(M^2-L^2)+2\omega^2LC-1][\omega^2CM-\omega^2CL+1]=0.$$

On solving this equation, there are two different values for ω^2, namely,

$$\omega^2=\frac{1}{C(L+M)}\quad\text{and}\quad\omega^2=\frac{1}{C(L-M)}.$$

Hence it is to be expected that before exact adjustment is made the crystal will oscillate with one frequency only, but when the length is exactly equal to the breadth, two new fundamental frequencies will appear.

EXPERIMENTAL RESULTS FOR RESONANCE WITH THE ELECTRICALLY COUPLED CIRCUITS

A perusal of the elastic properties will show that the beams are ideal for testing the resonance theory. The piezo-electric theory requires that they shall be cut with the orientations shown in fig. 1. Then on applying the elastic theory it is seen, for this special orientation only, that the elastic properties are exactly identical for the length and for the breadth, which are the directions along which the crystal may oscillate. An experimental test was made with a square beam of type C (fig. 1), of side 19·329 mm. This beam was chosen because a fairly large specimen was available. The difficulty of the test will be realized when it is pointed out that the frequency of the oscillation along the length or along the breadth, for a square beam of this size is about 86,000 c./sec. Before resonance can occur, it is probable that the frequency for the two directions must not differ by more than, say ten cycles or even a much smaller number. The method of procedure was to start with a square beam and reduce the length in steps of about one-ten-thousandth of an inch. After every adjustment the crystal was made to oscillate and one frequency only was found. Then the breadth was reduced in a similar manner for there was no method of determining whether one was approaching or receding from the required condition. A few score of adjustments were thus made with the length and breadth. When resonance occurred the crystal oscillated in a more vigorous manner causing the plates which were pressing lightly on the crystal and supplying the alternating field (henceforth called the terminals), to emit a high pitched audible note. Two new oscillations appeared of wave-lengths 3835 and 3190 m., together with the fundamental of wave-length 3487 m. The nature of the readings on the wavemeter, showed without any doubt, that they were new fundamental frequencies, and not overtones supplied by the wavemeter. The two new frequencies disappeared at the same time after a few minutes. It had previously been found that the wave-length increases with rise of temperature, both along the length and along the breadth, for beams of type *C*, and from the symmetry of the crystal the coefficients of expansion with temperature should be of equal amount in the two directions. The temperature of the crystal when oscillating rises a small fraction of a degree above the temperature of the surroundings. The disappearance of the oscillations was attributed to a slight change due to unequal heating. On breaking the circuit and allowing the crystal to cool for a short time the oscillations reappeared. The method was also tried of moving the crystal a short distance between the terminals in order to change the frequencies by a cycle or two, due to exceedingly small changes in the air-gap. This method also proved successful.

In some of the earlier tests it was found that the wave-lengths of the new frequencies drifted slightly. This may have been due to the damping caused by the vibrating terminals, which would have the same effect as the addition of a resistance to the coupled electrical circuits. This can also be explained by the fact that the natural frequencies of the two modes of vibration may not have been absolutely equal. A difference of a few cycles would alter the values of the new frequencies somewhat, but no means were available of determining at what instant the two natural frequencies were identical.

The experiment was repeated some months later, with the same crystal slightly reduced in size. For a square beam of side 19·218 mm. the fundamental wave-length was 3472 m. When resonance occurred, the longer wave-length was 3818 m., so that both values decreased by about 16 m. Both resonant frequencies disappeared before the shorter wave-length could be measured. This illustrates the extremely critical nature of the adjustment required for resonance. In this case the resonant condition was probably obtained by a slight temporary change of temperature due to the handling of the crystal specimen. It may perhaps, also be stated that more than two hundred adjustments were required to obtain the above results.

From the numerical results recorded, it may be calculated that the Coefficient of Coupling for the two equivalent electrically coupled circuits is 18 %.

It would appear that this resonance effect, requiring such critical adjustment, differs entirely from any previously shown.

Other interesting results were obtained with beams of type C. Broad beams of small thickness were employed and the wave-length was measured, while the length and thickness were kept constant and the breadth was gradually decreased. Since the beams used were of different dimensions a perusal of the table of results does not readily furnish much information nor does a graphical representation showing the relation between the wave-length and breadth It is much more instructive to plot the actual wave-length in metres divided by the lengths of the beams in millimetres, as ordinates, and the breadths as abscissae, the length and thickness being kept constant for each beam.

The results obtained for two beams only, of type *C*, are shown in graph 1, although the same effect was obtained with six different beams. On decreasing the breadth, it is seen that the wave-length decreases and a close examination shows that the curvature changes sign. When the breadth was reduced to about 0·60 of the length the wave-length suddenly increased by about 160 in 3000 m., and on further decreasing the breadth, a new curve was obtained Experiments with dust-figures seemed to prove without doubt that the beam was oscillating in a longitudinal manner along the length for all breadths, although, for the very broad beams, the dust figures were not very sharply defined. In the table (*C* 3) readings are given for one beam where both wave-lengths were obtained separately for the critical breadth of the beam, the breadth being 10·965 mm. and the length 18·314 mm., so that the ratio of breadth to length was 0·60. These discontinuities in the curves seem to be similar to those obtained by Lack and Hitchcock, when using quartz. If the results shown in the present paper can be attributed to resonance effects, then they seem to differ considerably from the results obtained with the square beams. With the latter two new frequencies appear at resonance. For the beams of decreasing breadth, single fundamental frequencies occur on both sides of the discontinuity. Still, it will be seen in the later part of the paper, that it is possible for oscillations along the length of a beam of length (l) and breadth ($0{\cdot}60l$), to have the same frequency as oscillations along the length of a beam of length ($0{\cdot}60l$) and breadth (l).

Experiments with changes in the lateral dimensions of the beams

Experimental results were obtained with the three beams shown in fig. 1. In the first part of the experiments the length and breadth were kept constant and the wave-lengths were measured on gradually reducing the thickness of the beam. In the second part of the experiment these thin broad beams were used and, keeping the length and thickness constant, the breadth was gradually cut down and the wave-length measured. For a long beam of very small lateral dimensions, the fundamental frequency of the longitudinal oscillations by this dynamical method should be determined by the simple equation:

$$f = \frac{1}{2l}\sqrt{\frac{Y}{\rho}}, \quad \text{or} \quad f = \frac{1}{2l}\sqrt{\frac{1}{E\rho}}, \tag{15}$$

where l is the length of the beam, ρ the density of its material, Y the Young's modulus and E the "Dehnungsmodul", a term used in crystal physics It should be pointed out, however, that this formula does not apply for beams of large lateral dimensions.

Rayleigh (1894) has estimated theoretically the error involved in neglecting the inertia of the parts of a vibrating rod of isotropic material which are not situated on the axis of the rod. Chree (1889) considers isotropic beams of rectangular cross-section and shows theoretically that any increase in the moment of inertia due to the lateral dimensions of the beam causes an increase of wave-length, whilst Davies (1933) has obtained a formula of corrections for crystals of Rochelle salt. It was found by experiment that Chree's formula subject to certain modifications, agreed best with the obtained results, if modified to the general form:

$$\lambda^2 - \lambda_0^2 = \lambda^2\left(k_1\frac{t^2}{l^2} + k_2\frac{b^2}{l^2}\right), \tag{16}$$

where λ is the observed wave-length for a beam of length l, breadth b and thickness t, λ_0 is the wave-length when b and t are very small, and k_1 and k_2 are constants varying as the square of the appropriate Poisson ratio, namely for the directions of length-thickness and of length-breadth respectively. These latter values can be obtained from the calculations already made for the lateral contractions accompanying an increase in length for the beams under test. It should be pointed out, however, that all the above correction formulae are based on the fact, that for an extension in

TABLE OF RESULTS

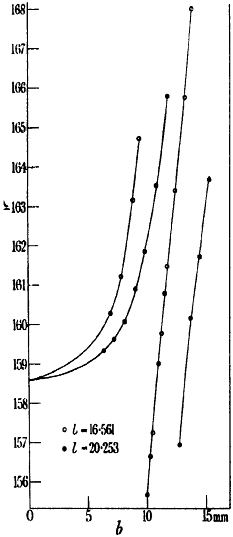
Beam *A* 1	*l* 16·573	*b* 13·868	*t*	9·474	8·214	6·860	5·725 mm.																		
			λ	4870	4869	4863	4861 m.																		
	l 16·573	*t* 5·725	*b*	13·868	12·804	11·791	10·874	9·906	8·539	7·468	6·383	5·067													
			λ	4861	4724	4597	4496	4412	4309	4247	4196	4146													
A 2	*l* 18·026	*b* 14·699	*t*	9·499	8·186	7·148	5·948																		
			λ	5226	5224	5223	5223																		
	l 18·026	*t* 5·948	*b*	14·699	13·766	12·974	12·114	11·049	8·994	7·783	6·350	5·851	4·719												
			λ	5223	5103	5014	4917	4817	4662	4602	4532	4495	4478												
A 3	*l* 19·857	*t* 5·932	*b*	17·203	15·206	14·591	13·508	11·582	10·230	7·694	6·655	5·738													
			λ	5897	5628	5554	5437	5266	5167	5033	4993	4959													
A 4	*l* 15·139	*b* 11·643	*t*	10·617	9·494	8·103	6·236	4·524																	
			λ	4300	4300	4300	4299	4293																	
A 5	*l* 18·435	*t* 5·944	*b*	15·172	14·452	13·812	13·650	13·398	12·674	11·988	11·364	10·940	10·439	10·203	9·989	9·753	9·423	9·182	8·971	8·713	8·163	7·577	7·049	6·287	5·605
	Experimental		λ	5359	5260	5187	5166	5129	5049	4984	4933	4895	4857	4838	4824	4807	4783	4769	4757	4745	4714	4677	4659	4632	4604
	Calculated from (17)		λ	5357	5258	5180	5160	5130	5052	4986	4931	4896	4857	4840	4825	4808	4787	4771	4759	4744	4713	4683	4659	4628	4603
	Difference			−2	−2	−7	−6	+1	+3	+2	−2	+1	0	+2	+1	+1	+4	+2	+2	−1	−1	+6	0	−4	−1
B 1	*l* 18·354	*b* 15·631	*t*	9·075	8·306	7·597	6·787	6·045	5·400	5·062															
			λ	2780	2775	2770	2768	2766	2764	2763															
	l 18·354	*t* 5·062	*b*	15·631	14·437	13·162	11·805	10·546	9·169	7·925	6·906	5·880	4·902												
			λ	2763	2722	2693	2673	2661	2650	2645	2641	2637	2634												
B 2	*l* 23·210	*t* 5·751	*b*	18·135	16·659	14·668	12·812	10·287	8·585																
			λ	3441	3406	3379	3362	3346	3337																
B 3	*l* 26·975	*t* 6·012	*b*	19·022	17·907	17·016	15·767	14·529	13·263	12·040	10·644	9·639	7·747	5·286											
			λ	3940	3929	3919	3908	3899	3890	3887	3878	3875	3871	3864											
C 1	*l* 16·561	*t* 5·207	*b*	13·823	13·196	12·438	11·754	11·557	11·245	11·005	10·490	10·307	9·962	9·410	8·839	7·846	6·960								
			λ	2782	2745	2706	2674	2663	2646	2634	2604	2594	2568	2728	2702	2670	2654								
C 2	*l* 20·253	*t* 5·512	*b*	15·304	14·481	13·698	12·761	11·732	10·835	9·855	9·042	8·153	7·273	6·370											
			λ	3315	3275	3234	3178	3358	3312	3278	3259	3242	3233	3227											
C 3	*l* 18·314	*b* 14·943	*t*	16·403	14·618	13·157	11·455	9·576	7·722	6·451	5·105														
			λ	3085	3076	3071	3067	3063	3059	3057	3055														
	l 18·314	*t* 5·105	*b*	14·943	13·805	12·398	11·714	11·257	11·112	10·965	10·965	9·984	8·900	7·785	6·828										
			λ	3055	2993	2918	2876	2843	2835	2819	3046	2980	2947	2914	2905										

length there are accompanying contractions along the breadth and along the thickness due solely to that extension. However, owing to the special piezo-electric properties of the beams used here, none of these formulae would be expected to apply. Consider beam C for instance, although the same reasoning applies to all of them, and suppose that a steady potential be applied in the direction OZ. Then the beam extends say, along the length and contracts along the breadth. The extension along the length causes a contraction along the thickness, and the contraction along the breadth an extension of equal magnitude along the thickness, so that for static fields there should be no change in the thickness of the beam and the correction term for the length-thickness directions is zero. For alternating fields experiments showed that the change in wave-length due to change in thickness was exceedingly small. The readings for the three types of beams are tabulated. It will be seen that, where measurements are made on reducing the thickness, there is a very small reduction of wave-length, in fact it is almost of negligible amount. For instance, with one of the beams (C 3) the thickness was reduced in steps from 16·403 to 5·105 mm. and the reduction in wave-length was from 3085 to 3055 m., a difference of about 1 % for a change in thickness of 11·3 mm. Similar results were obtained for all the beams so that the correction term $(k_1 t^2/l^2)$ may be omitted for thin beams. With regard to the second term $(k_2 b^2/l^2)$, the application of the static field produces an extension along the length which causes a contraction along the breadth as the correction formulae require, but there is now an added complication due to the fact that a further contraction takes place along the breadth due to the electric field. Hence it would be expected that for thin beams, on keeping the length constant and gradually reducing the breadth, the formula

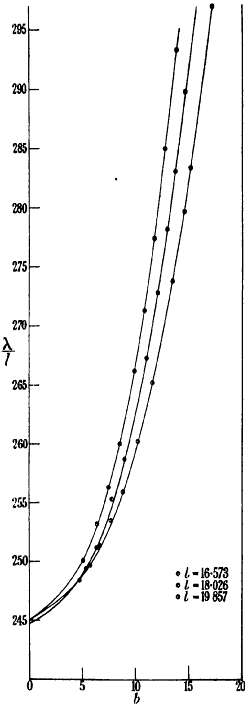
$$\lambda^2 - \lambda_0^2 = \lambda^2\left(k_2 \frac{b^2}{l^2}\right)$$

should apply, but k_2 for the different beams would not be expected to vary as the square of the Poisson's ratio.

Graph 2 shows the results obtained for beams A of fig. 1, with their lengths and breadths in the ZOX plane. Comparatively broad thin beams were used. While the length and thickness were kept constant the breadth was gradually decreased and the wave-length measured. From the theoretical calculations for a centimetre cube, for an extension along the length of $94{\cdot}2 \times 10^{-10}$ cm. the lateral contraction along the direction of thickness is $6{\cdot}7 \times 10^{-10}$ cm. and along the direction of breadth $82{\cdot}2 \times 10^{-10}$ cm. Poisson's ratio for the length-breadth direction (0·87) is therefore



GRAPH 1



GRAPH 2

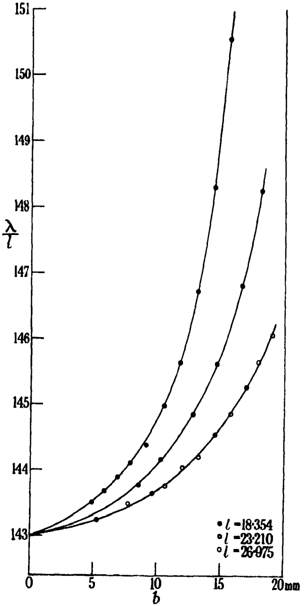
exceedingly large. The readings for beams A are tabulated and the results for three of them are shown in Graph 2. Since the three curves do not coincide it follows that the wave-length per millimetre length is a function of both the length and the breadth. Furthermore, for the beams of lengths 16·573, 18·026 and 19·857 mm , on reducing the breadths to a very small value the wave-lengths per millimetre length taken from the graph are 245·0, 244·7 and 245·0 m. respectively, while for beam (A 5), which is not shown on the graph, the value is 244·7. It is of interest to compare these values with those obtained on substituting the values found by the statical experiment equation (4) in (15). This calculated value is 234·1 m./mm. length, as compared with 245·0 m. by the present dynamical method, a difference of rather less than 5 %.

Graph 4 is drawn with $(\lambda^2-\lambda_0^2)/\lambda^2$ as ordinates and b^2/l^2 as abscissae. For the three beams shown in graph 2 it is found that all the points lie closely to a straight line of slope 0·430, which therefore gives the value of k_2. It is of interest to note that the formula applies over a very wide range, for in some of the broader beams the ratio of breadth to length is between 0·8 and 0·9 The results obtained with beam (A 5) are not plotted on the graph. Instead, two more rows are added to Table (A 5), one giving the wave-length as calculated by the formula

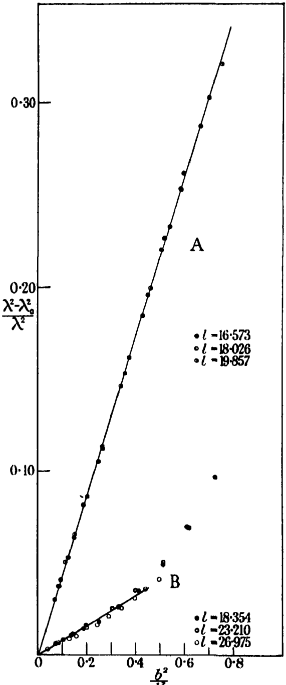
$$\lambda^2-(4510)=\lambda^2\left(0{\cdot}430\,\frac{b^2}{l^2}\right), \tag{17}$$

where 4510 is obtained by multiplying the length of the beam in millimetres by 244·7 The lower row gives the difference between the measured and calculated wave-length With this particular beam many readings were made to discover any small discontinuities due to resonance, and great care was taken that the temperature remained constant throughout the experiment. It will be seen that the calculated and experimental values agree very closely The wave-length increases by nearly 22 % for the broadest beam of (A 3) as compared with the infinitely narrow beam, and it may be noticed that this exceedingly large increase is associated with the abnormally large value of the Poisson's ratio for the length-breadth direction of the beam

The experimental results obtained with the beams B of fig 1 are tabulated and are shown in graph 3. A unit tensile stress, 1 g weight/sq. cm., along the length of a centimetre cube, gives an extension along the length of $37{\cdot}5\times10^{-10}$ cm., a contraction along the breadth of $5{\cdot}3\times10^{-10}$ cm., and a contraction along the thickness of $21{\cdot}05\times10^{-10}$ cm. It will be seen from the tabulated results that the change of wave-length on decreasing the thickness is again small in spite of the large value of the lateral contraction in this direction. The Poisson's ratio for the length-breadth direction is small, namely, 0·14, or about one-sixth of the corresponding value for the *A* beams. In graph 3 three thin beams are used of lengths 18·354, 23·210 and 26·975 mm. and the results show that the percentage increase in wave-



Graph 3


0·30
0·20
0·10
$\frac{\lambda^2-\lambda_0^2}{\lambda^2}$
A
B
● l = 16·573
○ l = 18·026
○ l = 19·857
● l = 18·354
○ l = 23·210
○ l = 26·975
0
0·2
0·4
0·6
0·8
$\frac{b^2}{l^2}$

GRAPH 4

length with increase of breadth is very much less than for the *A* beams. Thus on reducing the breadth of the *B* beams the maximum percentage change of wave-length is rather more than 3 %, while for the *A* beams of similar dimensions the change is about six times as much, so that the percentage changes of wave-length vary almost directly as the two Poisson's ratios for the length-breadth directions for the two sets of beams. Graph 3 shows that on reducing the breadth to an infinitely small value, the wave-length per millimetre length is 143·0 m. for all three beams. On substituting values found by the statical method, equations (3) and (15), the wave-length per millimetre length is 147·8 m. by calculation. Thus the wave-lengths agree to within about 3 % by the two methods for these beams. The lower part of graph 4 is obtained by plotting $(\lambda^2-\lambda_0^2)/\lambda^2$ as ordinates and b^2/l^2 as abscissae. The points lie fairly closely to a straight line, up to a point where the ratio of breadth to length is about two-thirds, but for broader beams the slope of the curve increases rapidly. The slope of the straight portion of the graph is 0·08 which therefore gives the value of k_2 for these beams.

The results obtained with the *C* beams have already been referred to. A unit tensile stress along the direction of length of a centimetre cube gives an extension of $47{\cdot}8\times10^{-10}$ cm., a contraction along the breadth of $10{\cdot}2\times10^{-10}$ cm. and a contraction along the thickness of $19{\cdot}25\times10^{-10}$ cm Owing to the anomalous nature of the results a curve is not plotted on graph 4 From graph 1 it is seen that long beams of small lateral dimensions have a wave-length of 158·6 m /mm. length as compared with 166 8 m. obtained from the statical experiment, equations (2) and (15), where broad beams were employed. There is thus a difference of about 5 % by the two methods. In comparing the values obtained for the three sets of beams by the dynamical and statical methods which differ by 3–5 %, it should be pointed out that broad beams were used in the statical experiments and infinitely long ones in the dynamical experiments. Furthermore no corrections are applied to the dynamical experiments for the "converse Piezo-electric effect". When a steady potential difference is applied to the crystal it extends along its length with the result that electric charges are developed due to this extension These produce stresses which prevent the crystal from extending to its full amount. This is equivalent to a decrease in the extension modulus so that the wave-length measured by the dynamical method is slightly too small. For a static field the decreases in the extension moduli vary as the square of the piezo-electric moduli which are large for the crystal beams under test. If this correction were applied then the values obtained by the statical and dynamical experiments for two of the sets of beams would be in still closer agreement.

The dynamical method is limited in its application for the determination of elastic constants since these are the only three differently oriented beams that can be employed for the production of longitudinal oscillations for this particular crystal, excepting the L-cut described by Cady (1937). In the statical experiment many difficulties arose since the beams were rather short, while corrections had to be applied for the depression of the supports and the "bite" of the knife-edges into the crystal face on adding the loads With one beam there was a depression of only ten interference fringes per load, so that it was necessary to read to one-tenth of a fringe, but the greatest difficulty arose owing to the fact that after applying the load a slight depression continued for a considerable time. In spite of these difficulties the results by the two methods are in good agreement.

The writer is pleased to acknowledge his indebtedness to the Trustees of the Dixon Fund for a grant for the purchase of apparatus.

His best thanks are due to Mr W. B. Medlam for advice on coupled circuits and to Dr Lownds for his kind interest in this work.

SUMMARY

Experiments are made with longitudinal oscillations in crystal beams of sodium ammonium seignette salt For a square beam a theory is deduced for the interaction of two longitudinal oscillations at right angles. These are considered as analogous to two coupled electrical circuits. If the natural frequencies in the two directions agree to within a small number of cycles per second, the theory requires that resonance shall occur with the production of two new fundamental frequencies This result is demonstrated experimentally.

Experiments are also made to determine the change in wave-length with change in breadth and thickness Theoretical calculations are made of the lateral contractions along the breadth and along the thickness, accompanying extensions along the length. The effect of thickness on the wave-length is found to be very small, in spite of the fact that the lateral contractions are large in two types of beams used. A change of breadth may change the wave-length by 20 %. The experimental results for two types of beams agree closely with an empirical formula, while one type of beam behaves in an anomalous manner.

REFERENCES

Cady 1937 *Proc. Phys. Soc.* **49**, 648.
Chree 1889 *Quart. J. Math.* **23**, 317-42.
Davies 1933 *Phil. Mag.* **16**, 104.
Lack 1929 *Proc. Inst. Radio Engrs, N.Y.*, **17**, 1123.
Hitchcock 1930 *Rev. Sci. Instrum.* **1**, 13-21.
Mandell 1928 *a* *Proc. Roy Soc.* A, **121**, 130.
— 1928 *b* *Proc. Roy. Soc.* A, **121**, 122.
Pierce 1923 *Proc. Amer. Acad. Arts Sci.* **59**, 82.
Rayleigh 1894 *Sound*, **1**, 251.
Voigt 1910 "Lehrbuch der Kristallphysik." Leipzig.

The alkaline permanganate oxidation of organic substances selected for their bearing upon the chemical constitution of coal

By R. B. Randall, Ph D., M. Benger, Ph.D , and C. M. Groocock, Ph D.

Of the Bone Research Laboratories

(*Communicated by Professor W. A. Bone, F.R.S.—Received* 3 *December* 1937)

In Part VIII of the previous papers upon the Chemistry of Coal from these laboratories (Bone, Parsons, Sapiro and Groocock 1935) it was shown that on oxidation by means of boiling alkaline permanganate the main organic substance of the lignin → peat → coal → anthracite series yields carbon dioxide, acetic, oxalic and benzene carboxylic acids, the proportions of the last named to the oxalic acid produced increasing with the maturity of the coal substance. Also, while every benzene carboxylic acid, except benzoic acid itself, was isolated from the oxidation products of some coal or other, in each and every case examined the penta- or hexacarboxylic acid predominated. Another notable result was that whereas no phthalic acids were detected among the benzenoid acids produced from lignins, and comparatively small proportions among those yielded by peats and brown coals, larger proportions thereof were yielded by bituminous coals and anthracites.

The results as a whole, while strongly supporting the view of the essential chemical continuity of the lignin → peat → coal → anthracite series, and of lignins rather than celluloses being its chief progenitors, suggested the possibility of the main coal substance having arisen through condensations of phenolic and amido-bodies with aldehydic bodies—much as systematic resins are now produced—and more recent observations have pointed to high pressures, rather than temperature, having been the principal factor in the natural maturing of the coal substance.

Although the essential benzenoid character of the main coal substance was thus proved, it has not yet been possible to deduce from the experimental evidence any constitutional formula for the "coal unit" in the same way as has been deduced for "cellulose" and "lignin" units. Such shortcoming, however, is due largely to lacunae in our systematic knowledge of the course and products of the alkaline oxidations of organic substances generally, for although such oxidations have been extensively employed in preliminary explorations of the molecular constitutions of individual organic substances, careful search of the literature has failed to disclose sufficient systematic knowledge for the detailed interpretation of the results of the alkaline permanganate oxidation of the coal substance in regard to its constitutional formula.

This being so, Professor Bone directed us to undertake a systematic qualitative and quantitative investigation of the alkaline permanganate oxidations of organic substances of known constitution selected for their bearing upon the coal problem, and the present paper embodies the results of the oxidations of some sixty such substances, each of which has been studied also in regard to the velocity of its oxidation. The last-named part of our task has led to the successful devising of a new experimental technique for the complete detailed determination of the oxidation-velocity curve for any particular organic substance, a matter which is reserved, however, for a further communication.

Previous observations upon the alkaline permanganate oxidations of organic substances generally are so scattered throughout the voluminous literature of organic chemistry that it is scarcely possible to summarize them briefly. The following may be cited, however, as bearing upon those of the particular classes of substances included in the present investigation:

Substance	Investigators	Acids obtained
Hexahydrofluorene	Pictet and Ramseyer (1911)	Phthalic, adipic and oxalic
Acetophenone $C_6H_5CO.CH_3$	McKenzie (1904)	Benzoylformic
Acenaphthene quinone	Charrier and Beretta (1924)	Naphthalic, hemimellitic, 2:6-dicarboxyphenylglyoxylic

Substance	Investigators	Acids obtained
Phenanthrene quinone	Charrier and Beretta (1924)	Diphenic
Acetomesitylene	Perkin and Tapley (1924)	Trimethylbenzoic, dimethylterephthalonic, 4 . 6-dimethylphthalonic, methylcarboxyphthalonic
β-Naphthoic acid	Ekstrand (1891)	Phthalic and trimellitic
Phenylacetic acid $C_6H_5CH_2\ COOH$	Prshevalski (1918)	Benzoylformic and benzoic
Phenylpropionic acid $C_6H_5CH_2\ CH_2\ COOH$	Prshevalski (1918)	Mandelic and benzoic
Phenylbutyric acid $C_6H_5CH_2CH_2CH_2COOH$	Prshevalski (1918)	Benzoylformic, benzoic, oxalic

Prshevalski (1918) showed that the attack upon an aromatic molecule begins on the carbon atom nearest the nucleus in a side chain with the formation of benzoic or an hydroxy-phenyl-acetic acid derivative, the rest of the side chain being simultaneously oxidized to carbonic anhydride or a dibasic acid. In our experiments it has been found that substitution of nuclear hydrogen by hydroxyl renders a benzenoid ring liable to disruption at such position under alkaline permanganate attack; that ketonic acids such as benzoylformic, phthalonic, and 2 . 6-dicarboxyphenylglyoxilic acids are remarkably resistant to such attack, and that variations in such conditions as temperature, alkali concentration, and the relative proportion of permanganate employed may have a qualitative as well as quantitative effect upon the products obtained We have also found, however, that by standardizing the conditions, so that they corresponded with those under which the previous coal oxidations were carried out, unvarying and easily reproducible results, as set forth in the accompanying tables, were obtainable in each case; and it should be understood that the results so recorded refer to such standardized conditions.

EXPERIMENTAL

(1) *Selection and classification of substances investigated*

The selection of substances for our investigations was based primarily upon their possible bearing upon the coal-constitutional problem, in view either of their relationship to possible progenitors of the coal substance, or of their chemical structures possibly containing, resembling, or having yielded these In regard to the last named, it seemed important to distinguish between those which are stable towards hot alkaline permanganate and those which may be readily further oxidized by it; because whereas, if formed at all during a coal oxidation, presumably the former would remain unchanged and accumulate throughout its course, the latter might undergo such rapid further oxidation as either to preclude or to curtail their survival among the end-products.

The substances actually investigated may be conveniently grouped as (1) carbohydrates and aliphatic carboxylic acids, (2) aromatic hydrocarbons, (3) aromatic substances containing —CO groups, and (4) aromatic carboxylic acid and ethers, and heterocyclic compounds. They were either purchased from reliable sources or synthesized in the laboratory, and in each and every case purity was established by analysis, melting point, and sometimes also by the preparation of some characteristic derivative. In tabulating and discussing the results each group will be considered separately.

(2) *Apparatus and procedure*

The experimental procedure has been based upon the "Carbon Balance Method" devised by Professor Bone and collaborators for determining the distribution of the carbon of the coal substance among its various alkaline-permanganate oxidation products, and combined with similar methods to that employed by them for the isolation and identification of those products. Determinations were always made (i) of the time required to reach the "end-point" of the oxidation under the standard experimental conditions, (ii) of the number of atoms of oxygen used per molecule of substance oxidized under the particular experimental conditions including relative concentrations of the reactants (e g. 3·5 g. substance, 5·6 g. caustic potash, x g. potassium permanganate added in 3·2 % solution, and 350 c c. water), (iii) of the individual oxidation products, and (iv) of the quantitative distribution between them of the carbon of the original substance. Finally, in all cases except that of the carbohydrates we were able to deduce empirical equations epitomizing the observed results of each oxidation.

With regard to the relative speeds of oxidation, it will be seen from the tabulated results that the times required for attainment of the end-point under like conditions varied from 20 min. to 166 hr. according to the substance investigated. No more than general significance should be attached, however, to the times given in the tables, especially in the cases of liquids or slightly soluble substances. Trimethylbenzoic acid, which was oxidized quantitatively to prehnitic acid in 25 hr., may be regarded as a convenient "standard" for comparative purposes.

Seeing that with the more highly resistant substances it was necessary to operate continuously over long periods of time, and that with liquid substances considerable difficulty from "bumping" was encountered, the apparatus described in Part VIII of previous papers from these laboratories had to be modified so that the reaction flask *A* (fig. 1) could be submerged in an oil bath which was electrically heated and controlled by means of the thermoregulator *B*. The contents of the flask were kept well stirred by means of the motor-driven ring stirrer *C* operating through the mercury seal *D*, and it was found necessary to fix a small condenser *E* on the stirrer shaft to prevent distillation from the flask to the mercury seal. The water trap *F* was made more efficient, the surrounding bath *G* being supplied with ice-cold water from a large reservoir (not shown). For the more rapid oxidations the electrical heating was replaced by gas heating.

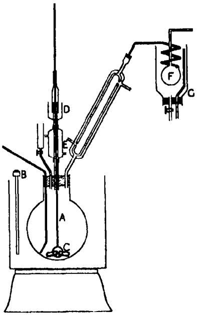


FIG 1

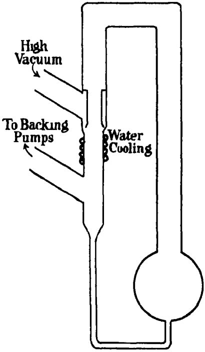


FIG. 2

The former esterification of the benzenoid acids resulting from these oxidations by the interaction of their silver salts with methyl iodide has been superseded by the use of diazomethane, and the technique for their fractionation under reduced pressure has been improved, using the higher vacua, of the order 10^{-6} mm. Hg, obtained by means of such oil-diffusion pumps as described by Hickman and Sandford (1930). Fig. 2 shows the particular form of pump employed by us.

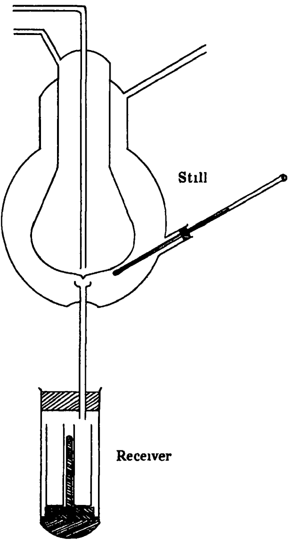


Fig. 3

Also, the use of the "vacuum sublimer" described in Part VIII of previous papers from these laboratories has been supplemented by that of a multiple receiver still (fig. 3) in such fractionations. This receiver had a small turntable carrying four tubes and fitted with a soft iron segment which allowed of its being rotated from outside by means of a magnet.

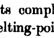
Some of the substances studied and/or maybe their oxidation products (e.g. fluorene and fluorenone) were so volatile in steam that part of them distilled into, and remained in, the condenser during the oxidation of the rest of the substance and were recovered unchanged at the end of the operations. In all such cases due allowance was made in respect of such recovered substances in calculating the "carbon balance" of the products.

Any benzoylformic and/or phthalonic acids occurring in the oxidation product could be estimated gravimetrically by precipitating their 2 : 4-dinitrophenylhydrazones from hot hydrochloric acid solution, the method having been proved accurate on its being tested for dilute solutions of the acids concerned In such connexion it should be noted that whereas the precipitation of the hydrazone of benzoylformic is effected almost immediately, that of the hydrazone of phthalonic acid only begins after an hour and requires four days for its completion The hydrazones were identified by analysis and mixed melting-point determinations.

TABULATION OF RESULTS

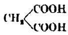
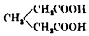
The detailed experimental results are presented in the series of four comprehensive tables (Nos. I to IV inclusive), each with its accompanying explanatory notes, incorporated herewith. All that need be added thereto are a few general observations in regard to the behaviour of the substances comprised in each table, and, at the end of the paper, a discussion of the implications of the results as a whole, as bearing upon the coal-constitutional problem. It is hoped that the results may also be useful in connexion with other similar constitutional problems.

Oxidations of (a) carbohydrates and (b) aliphatic acids

The substances investigated under (*a*) included ethylene glycol and polyglyoxal as well as representative carbohydrates, while those under (*b*) comprised glycollic, tartaric, malonic, succinic, glutaric and adipic acids. The polyglyoxal was a polymerized  $\begin{array}{l}\mathrm{H.C.O}\\ \quad\vdots\\ \mathrm{H.C\ O}\end{array}$ with properties as described by Harries and Temme (1907), while the cellulose I and II were from Swedish filter-paper and purified cotton-linters respectively.

The oxidations of polyglyoxal, glucose, glycollic and tartaric acids were all ended within about 20–25 min., and that of ethylene glycol in 45 min.· the ease of oxidation of the carbohydrates enumerated diminished in the

TABLE I. CARBOHYDRATES AND ALIPHATIC ACIDS

Substance	Formula	$KMnO_4$ ratio	Atoms of oxygen per mol.	Time to "end-point" of oxidation	% C appearing as			Total C accounted for
					CO_2	$(COOH)_2$	Acetic	
Ethylene glycol	CH_2OH CH_2OH	7·73	4·5	45 min.	63·9	32·8	Nil	96·7
Polyglyoxal	H C : O H.C : O	3·78	2·1	20 min.	48·0	47·8	2·5	98·3
Glucose	⌐CHOH $O(CHOH)_3$ ˡCH CH_2OH	5·58	9·5	20 min.	60·3	27·1	10·3	97·7
α-Methyl-*d*-glucoside	CH_2OH CHOH ˡCH $O(CHOH)_2$ └CH OCH_3	6·15	11·3	10 hr.	63·4	29·7	5·9	99·0
Maltose	$C_{12}H_{24}O_{12}$	5·56	9·5	2¼ hr.	59·7	29·1	10·1	98·9
Starch	$(C_6H_{10}O_5)_n$	6·08		3 hr.	66·0	24·1	7·5	97·6
Cellulose: I	—	6·54		–	49·4	46·3	2·9	98·6
II	–	6·19		—	54·2	43·6	2·2	100·0
Glycollic acid	CH_2OH COOH	3·35	2·4	20 min.	35·9	64·1	Nil	100·0
Tartaric acid:								
I	CHOH COOH	2·70	3·8 (5)	25 min	44·7	55·7	Nil	100·4
II	CHOH COOH	2·69	3·8	10 min.	35·5	62·5	Nil	98·0
Malonic acid	 $CH_2<^{COOH}_{COOH}$	3·33	3·3	14 hr	53·2	44·9	Nil	98·1
Succinic acid	CH_2COOH CH_2COOH	5·37	6·0	60 hr.	51·3	47·3	Nil	98·6
Glutaric acid	 $CH_2<^{CH_2COOH}_{CH_2COOH}$	7·18	9·0	75 hr.	57·7	41·1	Nil	98·8
Adipic acid	CH_2CH_2COOH CH_2CH_2COOH	8·50	11·8	115 hr.	58·6	41·8	Nil	100·4

order glucose → maltose → starch → methylglucoside → cellulose. The dibasic aliphatic acids, except tartaric with two hydroxyl groups, were all much more resistant to oxidation and in increasing degree as the series was ascended; and while all the malonic and the succinic acid was oxidized during the times (14 and 60 hr. respectively) stated in Table I, no less than 22·5 % of the glutaric and 38 % of the adipic acid remained unoxidized after 75 and 115 hr. respectively. Some darkening, due to slight resinification, occurred when the glucose, maltose, α-methyl-*d*-glucoside, and starch were first boiled with the aqueous caustic alkali prior to the addition of the permanganate.

In all cases substantially the whole of the carbon in the substance actually oxidized was accounted for as carbonic anhydride, oxalic and acetic acids, the last named being yielded by the polyglyoxal and the carbohydrates only—and in largest proportion by glucose and maltose—but not at all by either ethylene glycol, glycollic acid, or by any of the dibasic acids. Apparently its formation was due to the action of the boiling alkali upon the polyglyoxal and the carbohydrates. As might be expected, no benzenoid acid was ever formed among the oxidation products of any of the substances comprised in Table I, a circumstance which disposes of the suggestion that conceivably some aromatic structure might have been developed by the prolonged action of the alkali upon them.

The oxidations of ethylene glycol, glyoxal, glycollic acid and cellulose closely approximated to the empirical equations

$$3\begin{array}{c}\mathrm{H_2.C.OH}\\ \cdot\\ \mathrm{H_2:C.OH}\end{array}+14\mathrm{O}=\begin{array}{c}\mathrm{COOH}\\ \cdot\\ \mathrm{COOH}\end{array}+4\mathrm{CO_2}+8\mathrm{H_2O}$$

$$2\begin{array}{c}\mathrm{H.C:O}\\ \cdot\\ \mathrm{H.C:O}\end{array}+5\mathrm{O}=\begin{array}{c}\mathrm{COOH}\\ \cdot\\ \mathrm{COOH}\end{array}+2\mathrm{CO_2}+\mathrm{H_2O}$$

$$3\begin{array}{c}\mathrm{CH_2OH}\\ \cdot\\ \mathrm{COOH}\end{array}+7\mathrm{O}=2\begin{array}{c}\mathrm{COOH}\\ \cdot\\ \mathrm{COOH}\end{array}+2\mathrm{CO_2}+4\mathrm{H_2O}$$

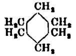
$$2\mathrm{C_6H_{10}O_5}+21\mathrm{O}=3\begin{array}{c}\mathrm{COOH}\\ \cdot\\ \mathrm{COOH}\end{array}+6\mathrm{CO_2}+7\mathrm{H_2O}$$

Oxidation of aromatic and cyclic hydrocarbons

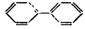
The following hydrocarbons were found to be so highly resistant to boiling alkaline permanganate that they may be regarded as practically unoxidizable by it, and accordingly they have been excluded from Table II, namely:

Benzene C_6H_6

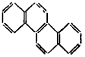
Hexahydrobenzene C_6H_{12}



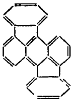
Diphenyl $C_{12}H_{10}$



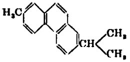
Chrysene $C_{18}H_{12}$



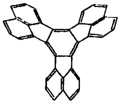
Rubicene $C_{26}H_{14}$



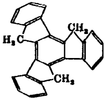
Retene $C_{18}H_{18}$



Decacyclene $C_{36}H_{18}$



Truxene $C_{27}H_{18}$



Trimethyltruxene $C_{30}H_{24}$

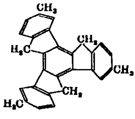
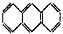
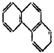
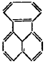
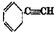
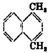
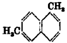
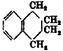
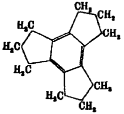
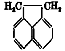
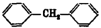
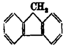
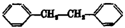
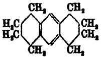
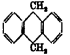
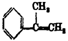


TABLE II. AROMATIC HYDROCARBONS

Substance	Formula	$KMnO_4$ ratio	Atoms of oxygen per mol	Time to "end-point" of oxidation	Products other than CO_2 and oxalic acid	% C appearing as				Equation
						CO_2	$(COOH)_2$	Benzenoid	Total	
Naphthalene		7 1	8 6	20½ hr	(*a*) Phthalonic acid (*b*) Phthalic acid	11·5	4·0	(*a*) 61 4 (*b*) 22 4	99·3	$3C_{10}H_8 + 24O = 2C_6H_4(COCOOH)(COOH)$ $+ C_6H_4(COOH)_2 + (COOH)_2 + 2CO_2$ $+ 2H_2O$
Anthracene		5 28	8 9	50 hr	(*a*) Phthalic acid (*b*) Phthalonic acid (*c*) Anthraquinone	11·3	5 6	(*a*) 17 2 (*b*) 3 2 (*c*) 61·0	98 3	$3C_{14}H_{10} + 23O = 2C_{14}H_8O_2 + 4CO_2$ $+ (COOH)_2 + C_6H_4(COOH)_2 + 3H_2O$
Phenanthrene		5 58	9·4	40 hr	(*a*) Diphenic acid (*b*) Phenanthrene-quinone	14 2	8 3	(*a*) 46·0 (*b*) 29 5	98 0	$20C_{14}H_{10} + 203O = 9C_{14}H_{10}O_4 + 6C_{14}H_8O_2$ $- 13(COOH)_2 + 44CO_2 + 15H_2O$
Fluoranthene		11 01	21 5	104 hr.	Mainly hemimellitic, some 2 6-carboxy-phenylglyoxilic acids	26 7	17·0	56·3	100·0	$2C_{16}H_{10} + 41O = 2C_6H_3(COOH)_3$ $+ 3(COOH)_2 + 8CO_2 + H_2O$
Phenyl-acetylene		5·09	5·0	61 hr.	(*a*) Benzoylformic acid (*b*) Benzoic acid	13 9	6·1	(*a*) 11·4 (*b*) 65 6	97·0	$8C_8H_6 + 44O = 6C_6H_5COOH + 2(COOH)_2$ $+ C_6H_5CO\ COOH + 10CO_2 + H_2O$
1 : 4-Dimethyl-naphthalene		9·45	14·0	33 hr	(*a*) Phthalic acid (*b*) Phthalonic acid (*c*) Mellophanic acid	20·2	4 9	(*a*) 1·0 (*b*) 21 8 (*c*) 44 9	92 8*	$20C_{12}H_{12} + 321O = 11C_6H_2(COOH)_4$ $+ 6C_6H_4(CO\ COOH)(COOH)$ $- 6(COOH)_2 + 3CH_3COOH + 58CO_2$ $+ 57H_2O$
1 : 6-Dimethyl-naphthalene		10·25	15·2	46 hr	Hemimellitic acid	10·1	19·7	65·5	95 3*	$10C_{12}H_{12} + 148O = 9C_6H_3(COOH)_3$ $+ CH_3COOH + 13(COOH)_2 + 11CO_2$ $+ 18H_2O$
Indene		6 1	6 7	25 hr.	(*a*) Phthalonic acid (*b*) Phthalic acid	9·7	3·2	(*a*) 54·2 (*b*) 27 7	94 8	$10C_9H_8 + 78O = 6C_6H_4(CO.COOH)(COOH)$ $+ 3C_6H_4(COOH)_2 + (COOH)_2 + 10CO_2$ $+ 12H_2O$
Tetrahydro-naphthalene		7 19	10·6	64 hr	(*a*) Phthalonic acid (*b*) Phthalic acid	12·2	2·9	(*a*) 54 3 (*b*) 27 7	97 1	$10C_{10}H_{12} + 103O = 4C_6H_4(COOH)_2$ $+ 6C_6H_4(CO\ COOH)(COOH)$ $+ (COOH)_2 + 12CO_2 + 29H_2O$

Tricyclo-trimethylene-benzene		10·69	20·1	16½ hr.	Mellitic acid	17·4	0·7	81·2	99·3	$C_{15}H_{18}+24O=C_6(COOH)_6+3CO_2$ $+6H_2O$
Acenaphthene		8·82	12·9	6 hr	Hemimellitic and 2·6-dicarboxyphenyl-glyoxilic acids	19·0	6·4	70·7	96·1	$2C_{12}H_{10}+26O=C_6H_3(COCOOH)(COOH)_2$ $+C_6H_3(COOH)_3+(COOH)_2+3CO_2$ $+3H_2O$
Diphenyl-methane		2·1	3·35	50½ hr	(*a*) Benzoic acid (*b*) Benzophenone	4·0	1·2	(*a*) 9·9 (*b*) 84·9	100·0	$20C_{13}H_{12}+69O=4C_6H_5COOH$ $+17C_{13}H_{10}O+(COOH)_2+9CO_2$ $+22H_2O$
Fluorene		2·83	4·5	45 hr.	(*a*) Phthalic acid (*b*) Phthalonic acid (*c*) Fluorenone	6·8	3·5	(*a*) 2·6 (*b*) 9·8 (*c*) 68·6	91·3	$12C_{13}H_{10}+76O=9C_{13}H_8O$ $+2C_6H_4(CO\cdot COOH)(COOH)$ $+3(COOH)_2+15CO_2+15H_2O$
Dibenzyl		8·2	14·2	130 hr	(*a*) Benzoic acid (*b*) Benzil	24·4	15·4	(*a*) 45·1 (*b*) 15·1	100·0	$7C_{14}H_{14}+102O=C_{14}H_{10}O_2$ $+7C_6H_5COOH+7(COOH)_2+21CO_2$ $+16H_2O$
Octohydro-anthracene		10·76	19·0	70 hr.	Pyromellitic acid	27·9	4·7	66·7	99·3	$C_{14}H_{18}+22O=C_6H_2(COOH)_4+4CO_2$ $+6H_2O$
9:10-Dihydro-anthracene		3·40	5·8	22 hr.	(*a*) Phthalic acid (*b*) Anthraquinone (*c*) Anthracene	3·3	0·8	(*a*) 7·1 (*b*) 65·9 (*c*) 16·8	93·9	$20C_{14}H_{12}+145O=3C_6H_4(COOH)_2$ $+13C_{14}H_8O_2+3C_{14}H_{10}+(COOH)_2$ $+30CO_2+43H_2O$
α-Methyl styrene		9·45	10·6	11½ hr.	(*a*) Benzoylformic acid (*b*) Benzoic acid	22·9	4·5	(*a*) 47·3 (*b*) 19·8	94·5	$20C_9H_{10}+205O=11C_6H_5CO\cdot COOH$ $+5C_6H_5COOH+4(COOH)_2+49CO_2$ $+48H_2O$

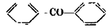
* Not including 2·3 and 2·4%, respectively, of the original carbon appearing as acetic acid in the products.

All the hydrocarbons included in Table II were oxidized slowly, yielding some aromatic acids or products in addition to carbonic anhydride, oxalic acid and water. Nearly 2·5 % of the carbon of the two dimethyl naphthalenes oxidized appeared as acetic acid in the product; but in no other case was acetic acid formed. Decahydronaphthalene yielded a large proportion of phthalonic acid. In each case practically the whole of the carbon of the substance oxidized was distributed among such products, and the oxidation could be approximately expressed by an empirical equation.

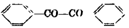
Oxidations of aromatic substances containing a =CO *group*

The following substances were found to be practically unoxidizable by the boiling alkaline permanganate

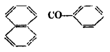
Benzophenone $C_{13}H_{10}O$



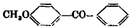
Benzil $C_{14}H_{10}O_2$



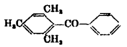
1-Benzonaphthone $C_{17}H_{12}O$



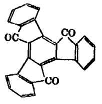
4-Methoxybenzophenone $C_{14}H_{12}O_2$



2 . 4 : 6-Trimethylbenzophenone $C_{16}H_{16}O$



Tribenzoylenebenzene $C_{27}H_{12}O_3$



From which it would seem as though =CO groups situated between two benzene or alkyl-benzene nuclei are usually immune from attack

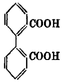
From Table III it will be seen that while acetophenone was readily oxidized, its 4-methyl derivative was much less so, on the other hand, 4-methylbenzophenone was more oxidizable than benzophenone. Such differences were probably, however, due to differences in the solubility of these substances. It will also be seen that in each case substantially the whole of the carbon of the substance oxidized could be accounted for as carbonic anhydride, oxalic acid, and some aromatic acid or compound in the products, and that its oxidation could be approximately expressed by an empirical equation. The following other points should be noted, namely (i) the extreme slowness of the oxidation of 4-methylbenzophenone and 2-methoxy-1-benzonaphthone, both of which contain a —CO group between two benzene nuclei; (ii) the marked acceleration of oxidation effected by the replacement by hydroxyl either of the methyl group of 4-methylbenzophenone or of the methoxyl group of 2-methoxy-1-benzonaphthone, and (iii) that about 2·3 % of the carbon of the 4-methylbenzophenone oxidized appeared as acetic acid in the products.

It may also be remarked that whereas Charrier and Ghigi (1933) obtained chiefly 2′ . 3-dicarboxy-2-diphenylglyoxilic acid, together with a small quantity of anthraquinone-1-carboxylic acid, by oxidizing benzanthrone with permanganate at 80–90° in a very high concentration of alkali, our products were 2 : 6-dicarboxyphenylglyoxilic acid and a small quantity of phthalic acid, the difference being probably due to the lower temperature and higher alkali concentration employed by them as compared with our conditions

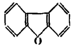
Oxidations of aromatic carboxylic acids, heterocyclic compounds, etc.

The following substances were found to be practically immune from oxidation by the boiling alkaline permanganate, namely

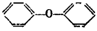
2 : 2′-Diphenic acid $C_{14}H_{10}O_4$



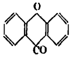
Diphenylene oxide $C_{12}H_8O$



Diphenyl ether $C_{12}H_{10}O$

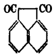
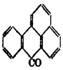
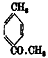
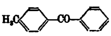


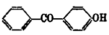
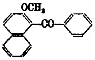
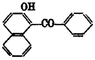
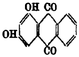
Xanthone $C_{13}H_8O_2$



From Table IV it will be seen, (i) that 2 : 4 . 6-trimethylbenzoic acid was oxidized quantitatively to the corresponding 1 : 2 . 3 : 5-benzene tetracarboxylic (prehnitic) acid in about 25 hr., each of its three methyls

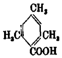
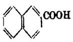
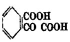
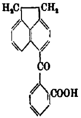
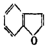
TABLE III. AROMATIC SUBSTANCES CONTAINING A =CO GROUP

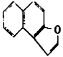
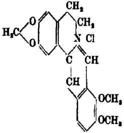
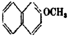
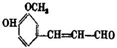
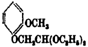
Substance	Formula	$KMnO_4$ ratio	Atoms of oxygen per mol.	Time to "end-point" of oxidation	Products other than CO_2 and oxalic acid	% C appearing as				Equation
						CO_2	$(COOH)_2$	Benzenoid	Total	
Acetophenone		3·02	3 4	22 min.	(a) Benzoylformic acid (b) Benzoic acid	6 7	Nil	(a) 58 2 (b) 33 5	98 4	$3C_8H_8O + 10O = 2C_6H_5CO\ COOH + C_6H_5COOH + CO_2 + 3H_2O$
Acenaphthene quinone		5·21	9·0	4½ hr.	Mainly 2.6-dicarboxyphenylglyoxilic acid	18 1	1 1	79·7	98 9	$C_{12}H_6O_2 + 9O = C_6H_3(COCOOH)(COOH)_2 - 2CO_2$
Benzanthrone		10·12	22 1	80 hr	Mainly 2.6-dicarboxyphenylglyoxilic, some phthalic, acid	20·3	18·0	61 3	99·6	$10C_{17}H_{10}O + 204O = C_6H_4(COOH)_2 + 9C_6H_3(CO\ COOH)(COOH)_2 + 17(COOH)_2 + 38CO_2 + 3H_2O$
4-Methyl-acetophenone		6·16	7 8	3 hr.	(a) Terephthalic acid (b) Terephthalonic acid	15 4	0·9	(a) 44 4 (b) 39 3	100·0	$9C_9H_{10}O + 74O = 4C_6H_4(COCOOH)(COOH) + 4C_6H_4(COOH)_2 + 13CO_2 + 21H_2O$
4-Methyl-benzophenone		3·81	7·3	160 hr.	Benzophenone-4-carboxylic acid	10·4	9·4	74·7	94 5*	$8C_{15}H_{12}O + 75O = 6C_{15}H_{10}O_3 + CH_3COOH + 5(COOH)_2 + 16CO_2 + 11H_2O$

4-Hydroxy-benzo-phenone		6·15	11·6	3 hr	(*a*) Benzoylformic acid (*b*) Benzoic acid	22·7	18·9	(*a*) 23·1 (*b*) 33·5	98·2	$10C_{13}H_{10}O_2 + 11O O = 4C_6H_5CO\ COOH + 7C_6H_5COOH + 11(COOH)_2 + 27CO_2 + 6H_2O$
2-Methoxy-1-benzo-naphthone		8·06	20·5	166 hr	(*a*) Phthalic acid (*b*) Phthalonic acid (*c*) Benzoic acid	27·2	14·0	(*a*) 2·1 (*b*) 17·2 (*c*) 36·1	96·6	$2C_{18}H_{14}O_2 + 34O = 2C_6H_5COOH + C_6H_4(CO\ COOH)(COOH) + 2(COOH)_2 + 9CO_2 + 3H_2O$
2-Hydroxy-1-benzo-naphthone		3·25	7·7	13 hr	(*a*) Benzoic acid (*b*) Phthalonic acid (*c*) Phthalic acid	13·3	3·6	(*a*) 36·1 (*b*) 5·9 (*c*) 36·3	95·2	$C_{17}H_{12}O_2 + 8O = C_6H_4(COOH)_2 + C_6H_5COOH + 2CO_2$
1 : 2-Dihy-droxyanthra-quinone		5·05	10·2	15 min.	(*a*) Phthalonic acid (*b*) Phthalic acid	34·5	6·7	(*a*) 52·7 (*b*) 4·1	98·0	$2C_{14}H_8O_4 + 23O = (COOH)_2 + 2C_6H_4(COCOOH)(COOH) + 8CO_2 + H_2O$

* Not including 2·3% of the original carbon appearing as acetic acid in the products.

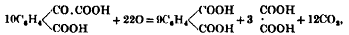
TABLE IV. AROMATIC CARBOXYLIC ACIDS, HETEROCYCLIC COMPOUNDS, ETHERS AND ALDEHYDES

Substance	Formula	$KMnO_4$ ratio	Atoms of oxygen per mol	Time to "end-point" of oxidation	Products other than CO_2 and oxalic acid	% C appearing as CO_2	% C appearing as $(COOH)_2$	% C appearing as Benzenoid	Total	Equation
2.4.6-Tri-methyl-benzoic acid		5·86	9·1	25½ hr.	Prehnitic acid	2·8	Nil	95·6	98·4	$C_{10}H_{12}O_2 + 9O = C_6H_2(COOH)_4 + 3H_2O$
2-Naphthoic acid		5·03	8·2	3½ hr.	(*a*) Phthalic acid (*b*) Phthalonic acid (*c*) Trimellitic acid	10·7	4·9	(*a*) 17·4 (*b*) 51·4 (*c*) 13·4	97·8	$10C_{11}H_8O_2 + 81O = 2C_6H_4(COOH)_2 + 6C_6H_4(CO\,COOH)(COOH) + 3(COOH)_2 + 2C_6H_3(COOH)_3 + 16CO_2 + 7H_2O$
Phthalonic acid		1·33	2·45	40 hr.	Phthalic acid	11·2	8·8	81·2	101·2	$10C_9H_6O_5 + 22O = 9C_6H_4(COOH)_2 + 3(COOH)_2 + 12CO_2$
Acenaphthoyl-benzoic acid		4·68	13·5	15 hr	Phthalic, phthalonic 2·6-dicarboxyphenyl-glyoxilic, mellophanic, and an unidentified acid	13·6	1·7	84·7	100·0	—
Coumarone		15·38	17·2	3 hr	None	86·6	11·9	Nil	98·5	$2C_8H_6O + 35O = (COOH)_2 + 14CO_2 + 5H_2O$

2 : 1-Naphtho-1′ : 2′-furane		7·04	11·25	11 hr	(*a*) Phthalonic acid (*b*) Phthalic acid	24·9	1·4	(*a*) 43·0 (*b*) 27·2	96·5	$10C_{12}H_8O + 113O = 4C_6H_4(COOH)_2$ $+ 6C_6H_4(CO.COOH)(COOH)$ $- (COOH)_2 + 32CO_2 + 9H_2O$
Berberine hydro-chloride		7·87	27·8	80 hr	Hemipinic and hydrastinic acids	38·3	26·6	32·8	97·7	$3C_{20}H_{17}NO_4\ HCl + 99O = 3HCl + 3HNO_3$ $+ (CH_2O_2)C_6H_2(COOH)_2$ $+ (CH_3O)_2C_6H_2(COOH)_2 + 8(COOH)_2$ $+ 25CO_2 + 8H_2O$
2-Naphthyl-methyl ether		7·04	10·5	7½ hr	(*a*) Phthalonic acid (*b*) Phthalic acid	15·9	2·2	(*a*) 64·5 (*b*) 11·3	93·9	$5C_{11}H_{10}O + 50O = C_6H_4(COOH)_2$ $+ 4C_6H_4(CO\ COOH)(COOH)$ $+ (COOH)_2 + 9CO_2 + 9H_2O$
Conifer-aldehyde		12·52	21·1	2½ hr	None	74·3	20·5	Nil	94·8	$C_{10}H_{10}O_3 + 21O = (COOH)_2 + 8CO_2$ $+ 4H_2O$
1-Methoxy-2ω-aldomethoxy-benzene di-ethyl acetal		12·41	28·2	2 hr	None	34·9	63·2	Nil	98·1	$C_{13}H_{20}O_4 + 28O = 4(COOH)_2$ $+ 5CO_2 + 6H_2O$

being oxidized to a carboxyl, COOH, group, the time required being comparable with that for a typical Shafton bituminous coal (Barnsley); (ii) that phthalonic acid was yielded by 2-naphthoic acid, 2 : 1-naphtho-1′ : 2′-furane, and 2-naphthyl methyl ether; and (iii) that coumarone, coniferaldehyde and 1-methoxy-2ω-aldomethoxybenzene diethyl acetal were all oxidized straightway to carbonic anhydride, oxalic acid and water without any appreciable formation of aromatic acids.

Attention is also directed to relatively greater stability of phthalonic as compared with the other carboxylic acids included in the table, although it was more oxidizable than either phthalic or the other benzene carboxylic acids examined. It yielded carbonic anhydride, oxalic and phthalic acids substantially in accordance with the equation:



$$10C_6H_4\left\langle\begin{matrix}CO.COOH\\ COOH\end{matrix}\right. + 22O = 9C_6H_4\left\langle\begin{matrix}COOH\\ COOH\end{matrix}\right. + 3\begin{matrix}COOH\\ \cdot\\ COOH\end{matrix} + 12CO_2,$$

although some 40 % remained unoxidized after the "end-point" of 40 hr.

Berberine hydrochloride was oxidized slowly, yielding carbonic anhydride, nitric, hydrochloric, oxalic, hemipinic and hydrastinic acids, the two last named being relatively about as resistant to oxidation as phthalonic acid.

Discussion

From our results, in conjunction with those of other workers, it would appear, in regard to oxidation by boiling alkaline permanganate, that.

(1) Unreduced and unsubstituted aromatic hydrocarbons containing more than three "fused" rings are practically unattacked, a circumstance possibly due to their insolubilities in the reagent.

(2) Alkyl-substituted aromatic and reduced cyclic hydrocarbons are, in general, more susceptible to oxidation, reduced (five- or six-membered) rings being completely oxidized to the corresponding carboxylic acids. Thus, for example, tetrahydronaphthalene, octohydroanthracene, and tricyclotrimethylenebenzene gave large yields of phthalic and phthalonic, pyromellitic, and mellitic acids respectively. Evidence was forthcoming of the formation of a benzene carboxylic acid by the oxidation of a fully reduced ring to a benzene ring, e g. in the oxidation of decahydronaphthalene.

(3) Unsaturated linkages in the side chain of an aromatic hydrocarbon provide a ready point of attack, e.g. in phenylacetylene and methylstyrene.

(4) A $=CH_2$ linked to a benzene ring is oxidized primarily to $=CO$,

e.g. dibenzyl to benzil, fluorene to fluorenone, and diphenylmethane to benzophenone.

(5) Substitution of hydrogen by hydroxyl in the nucleus of an aromatic substance renders it very susceptible to oxidation. Thus, for example, 4-hydroxybenzophenone, 1 · 2-dihydroxyanthraquinone, and coniferaldehyde were among the most rapidly oxidizable substances investigated, their rings being primarily attacked and disrupted at the point of attachment of the hydroxyl group to the benzenoid nucleus.

(6) Similarly, the attachment of a methoxy group to the nucleus of an aromatic substance increases its vulnerability to the oxygen attack. Thus, for example, methoxybenzonaphthone was much more readily oxidized than benzonaphthone.

(7) On the contrary, a —CO group attached to an aromatic nucleus stabilizes it against the oxygen attack. Thus, for example, benzophenone, benzil, anthraquinone, and fluorenone were practically unoxidizable, and although acetophenone was readily oxidized it yielded benzoyl-formic acid, in which the —CO group was retained, as its main product.

(8) Unsubstituted aromatic carboxylic acids were no more rapidly oxidized than the corresponding hydrocarbons, and generally speaking were stable, e g. all the benzene carboxylic, diphenyl carboxylic and benzophenone carboxylic acids.

(9) The side chains of an alkyl-benzene carboxylic acid are, however, readily oxidized, e g. 2 . 4 . 6-trimethyl benzoic acid was quantitatively oxidized to prehnitic acid, which was stable.

(10) In no case investigated was the structure of any oxidation product more complex than that of the original substance oxidized; in other words the oxidations were all degradations. In some cases evidence was forthcoming of an oxidation proceeding simultaneously in different ways, e.g. that of acenaphthoylbenzoic acid which yielded simultaneously phthalic, 2 . 6-dicarboxyphenylglyoxilic and mellophanic acids.

In regard to the bearing of our investigation upon the chemical constitution of coal and allied substances, it would seem as though the various formulae hitherto put forward for lignin and humic acids are inadequate inasmuch as such formulae do not contain structures capable of yielding the alkaline permanganate oxidation products which have, in fact, been obtained from them. The structures possibly present in coals would seem to be unreduced benzene rings linked through side chains or oxygen heterocyclic rings. Fused reduced rings are probably not present to any great extent, seeing that ketonic acids have often been sought, but never found, among the products of coal oxidations.

In conclusion the authors desire to thank Professor Bone for having suggested and directed the investigation and the Department of Scientific and Industrial Research for grants which have enabled it to be carried out.

SUMMARY

The paper describes the results of systematic qualitative and quantitative investigations of some sixty organic substances, namely, (i) carbohydrates and aliphatic carboxylic acids, (ii) aromatic hydrocarbons, (iii) aromatic substances containing —CO groups, (iv) aromatic carboxylic acids and ethers, and (v) heterocyclic compounds, selected for their bearing on the coal-constitutional problem, and discusses the principal conclusions to be drawn therefrom.

REFERENCES

Bone, Parsons, Sapiro and Groocock 1935 *Proc. Roy. Soc.* A, **148**, 492.
Charrier and Borotta 1924 *Gazz. chim. ital* **54**, 988.
Charrier and Ghigi 1933 *Gazz. chim. ital.* **63**, 685.
Ekstrand 1891 *J. prakt. Chem.* (2) **43**, 427.
Harries and Tomme 1907 *Ber dtsch. chem. Ges* **40**, 165.
Hickman and Sandford 1930 *Rev. Sci. Instrum.* **1**, 140.
McKenzie 1904 *J. Chem Soc* p. 1254.
Perkin and Tapley 1924 *J. Chem. Soc.* p 2428.
Pictet and Ramseyer 1911 *Ber dtsch. chem. Ges* **44**, 2486.
Prshevalski 1918 *J Soc. phys -chim Russe*, **49**, 567

ERRATUM

Proc. Roy. Soc., *A*, *vol.* 165, *p.* 310, *l.* 18

for 10^{-4} *read* 10^{-5}

Photo-electric measurements of the seasonal ~~variations~~ in daylight around $0{\cdot}41\mu$, from 1930 to 1937

By W. R. G. Atkins, Sc.D., F.R.S.

Head of the Department of General Physiology, Marine Biological Laboratory, Plymouth

(*Received* 1 *February* 1938)

An account has already been given of the standardization of photo-electric cells for the measurement of daylight (Poole and Atkins 1935) and of the recording of daylight using a Burt vacuum sodium cell and Cambridge thread recorder (Atkins and Poole 1936). It was shown that the measurements relate to the ultra-violet, violet and blue, but are chiefly an indication of the changes in light of wave-length about $0{\cdot}41\,\mu$, a region in which the eye is of very low sensitivity.

CONSTANCY OF THE CELL USED

The results published were for the year 1930. Since the measurements now given are a continuation of these it is obvious that their value depends upon the constancy of the cell over this long period, eight years.

Evidence based on standardizations may be found in the 1936 paper, proving constancy from 1930 till May 1934. During this period the opal-flashed diffusing glass (opal) had been renewed once, when blown off and smashed on the night of 19 September 1930. The new disk was cut from the same sheet of glass as the old, and such disks were usually uniform in transmission to 1 or 2 %. Furthermore, examination of the records before and after the change showed no sign of any alteration occasioned thereby. Though certain of the results for 1930 are so high as to lead one to suspect a subsequent loss of sensitivity, yet there is much internal evidence of constancy in the tables of results. For example, the mean illumination integral, in kilo-lux hours, was 315 for March 1930 and 319 for March 1937. The readings shown in Table I also negative the idea that there has been any decrease in the sensitivity of the cell.

No error of any importance was introduced by variations in anode potential or by the minor alterations in the levelling of the opal; this was cleaned daily and the space below it was wiped dry.

TABLE I. DAILY MAXIMUM CURRENT IN MICROAMPERES, FROM BURT SODIUM CELL NO. 299, ON FRONT PARAPET OF LABORATORY ROOF. BRIGHT DAYS IN 1933 WERE COMPARED WITH THE SAME DATES IN OTHER YEARS

	1930	1933	1937
18 Mar.	23 85	23·90	24·10
21	23·90	22 75	10·70
22	22 60	22·40	23·90
23	9 60	23 50	23·00
24	22·85	21·00	25·30
30	26 00	26·70	26·70

THE RECORDING OF VERTICAL ILLUMINATION

Daily charts were obtained and were all measured by the author exactly as previously described for 1930 Defects in the high-tension ignition rubber-coated cable became more numerous in time, so the means for 1932 and 1933 are somewhat less accurate than for the other years A day lost was always a wet (namely dark) one, so in order not to raise the average unduly, a typical value for a wet day at the same season was inserted. The possibility of error from electrical leaks was guarded against by frequent determinations of the zero. The recorder was out of action for over two months in the last quarter of 1933, while being restandardized and having the new lead alloy sheathed cable placed in position. The photometer case was not opened, but the cable was cut off near the case and the new cable was attached to the old, the remains of which were heavily coated with a bituminous insulating compound. As trouble arose later on at the junction, this was housed in a metal box after adequate coating.

VERTICAL ILLUMINATION THROUGHOUT THE YEARS 1930–7

In the tables which follow the monthly maxima in the vertical columns are shown in heavy type and the minima in italics. Yearly maxima have an asterisk, but in later tables are in heavy type. In view of the occurrence of values during 1930 (Atkins and Poole 1936, Tables III and IV), which subsequent records showed to be remarkably high, the charts were examined again to make sure that such high figures were not due to electrical leaks which had escaped notice before. The local meteorological returns were also consulted as to the occurrence of rain. There remains no doubt that these values for 1930 are correct, and in a certain number of them additional observations, noted at the time, give reason to believe that the day was particularly bright. Thus on 22 April 1930, the ratio of total (vertical) illumination to diffuse illumination, denoted as β, was as high as 3·25 at 1·5 p.m. 7 July 1930 shows the maximum illumination integral for the whole eight years, namely, 1323 kilolux hours, notes record that the wind was north to north-west, with air exceptionally clear, sky very blue, a few clouds at noon, none in the afternoon, at 12.5, 2.25 and 4.0 p.m. G.M.T. β was 4·25, 3·44 and 3·50 respectively. There is no doubt about the genuineness of a few outstanding records, but this does not explain why certain months gave exceptionally high records irrespective of the sunshine or rain.

The measurements were undertaken in the first place in connexion with the study of the seasonal changes in the plankton in the English Channel. Daily values of the maximum vertical illumination in kilolux and of the total vertical illumination in kilolux hours have accordingly been filed for use at Plymouth. But in the tables which follow only the mean, greatest and least values are given for the daily maximum for each month, and the corresponding results for the illumination integral. There are thus no published tables corresponding to No IV of Atkins and Poole (1936).

Table II shows that the daily maximum vertical illumination was least in November 1934 with 3·2 kilolux (see Atkins and Poole 1936, Plate 27 D) and greatest in July 1930 with 197·6 kl., a range of almost sixty-two fold. During 1930, April, May, July and August, gave maxima over 190 kl. outstandingly high values, with a mean maximum of 150·9 for July. High means were also obtained for April, May and August. The June value too, 119·3 kl., was the highest for any June though surpassed in May 1932 which gave 126·2 kl. Though the mean maxima were quite ordinary values for January, February and March 1930, yet from April to November they were exceptionally high, December was only surpassed slightly by 1935 and January and February 1931 were again high, with March quite a normal value. Occasional high maxima may be explained by exceptional conditions, such as the reflexion from towering white clouds before a snow-storm on 28 February 1931. Anything that tends to raise the average altitude of the light naturally raises the vertical component. Thus on 21 December 1937, there was a diffuse misty grey sky, after extremely heavy rain the previous night, and the sun broke through, giving for a short time $V = 21{\cdot}4$ kl., as against maxima for the 21st on other years as follows: 1929, 17·2 kl.; 1930, 19·9 kl.; 1931, 16·0 kl.; 1932, 17·4 kl.; 1934, 17·9 kl., but for 22nd; 1935, 17·5 kl. and 1936, 4·9 kl. Here again we have no indication of any loss of sensitivity. But when we turn to the mean

TABLE II. DAILY MAXIMUM VERTICAL ILLUMINATION IN KILOLUX, AT PLYMOUTH, LATITUDE 51·5 N. TOP SECTION, LEAST VALUES FOR EACH MONTH, MIDDLE SECTION, GREATEST VALUES AND BOTTOM SECTION, MONTHLY MEAN OF DAILY MAXIMA YEARLY MAXIMA HAVE ASTERISK

Year	Jan.	Feb	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1930	8·7	11·8	24·9	53·2	*38·0*	41·8	64·6	77·1	47·5	20·9	8·6	6·9
1931	11·8	19·5	17·7	46·0	51·7	25·8	48·6	48·3	*19·0*	22·6	10·6	*3·8*
1932	7·6	*4·8*	27·2	32·3	57·7	40·3	52·1	34·2	21·3	11·4	9·9	4·4
1933	11·2	15·2	36·1	57·0	74·1	47·5	53·6	52·8	20·9	—	—	—
1934	5·1	12·5	25·2	41·8	44·5	41·8	83·6	23·9	24·5	10·4	*3·2*	4·7
1935	6·1	19·5	*16·1*	43·7	60·0	*24·7*	41·8	29·3	32·3	12·0	9·5	6·7
1936	*4·4*	8·6	27·4	*28·1*	85·5	56·6	*26·9*	*19·2*	36·1	*8·4*	7·8	4·9
1937	5·0	7·6	16·3	34·6	54·7	61·6	34·6	42·9	22·8	10·1	4·6	6·8
1930	35·1	43·7	98·8	196·1	197·6*	174·8	195·3	193·0	130·7	98·4	79·8	31·2
1931	39·5	82·5†	98·4	151·2	150·9	152·8*	134·5	131·1	107·5	72·2	44·6	22·6
1932	34·2	56·6	130·0‡	145·2	168·7*	139·1	134·9	105·6	92·3	73·7	44·1	23·7
1933	38·8	70·1	101·8	159·6*§	136·8	127·3	107·2	91·2	76·7	—	—	—
1934	34·3	68·4	114·0	153·5	162·6*	137·6	116·7	119·7	116·7	75·2	51·3	23·8
1935	37·3	70·9	94·6	135·3	160·0*	153·9	118·2	95·0	106·4	88·9	54·0	31·9
1936	39·5	66·9	98·9	150·1	146·7	156·6	168·3*	119·7	93·1	66·1	42·2	27·7
1937	32·7	66·5	101·8	119·3	143·6*	134·2	129·2	104·1	89·7	61·6	42·9	27·4
1930	21·3	*32·9*	68·8	142·2	149·0	119·3	150·9*	138·3	98·0	59·3	31·1	18·1
1931	26·0	44·1	69·9	104·1	114·4*	107·0	108·5	97·3	67·0	46·0	27·6	*12·4*
1932	19·3	33·8	71·8	111·0	126·2*	100·7	93·9	72·2	55·1	48·2	*19·4*	16·6
1933	24·0	39·9	68·9	96·2	108·7*	*94·5*	*86·3*	*71·7*	*53·9*	—	—	—
1934	*17·2*	40·6	77·0	101·5	117·4*	108·8	100·2	92·7	73·3	47·1	21·7	14·9
1935	17·6	38·9	*64·2*	99·2	107·2*	105·3	90·4	71·8	77·1	47·1	28·7	18·6
1936	21·5	40·3	67·6	100·7	109·1	110·4*	101·5	84·7	67·6	46·0	22·9	15·7
1937	18·3	35·9	75·2	*85·1*	*100·7*	111·8*	95·8	79·9	69·8	*44·5*	24·4	17·9

† Feb. 28, see Plate 24 D, Atkins and Poole, 1936. ‡ Mar. 31. § April 30.

TABLE III. VERTICAL ILLUMINATION INTEGRAL, IN KILOLUX HOURS A DAY, AT PLYMOUTH. TOP SECTION, LEAST VALUES FOR EACH MONTH, MIDDLE SECTION, GREATEST VALUES AND BOTTOM SECTION MEAN MONTHLY VALUES. YEARLY MAXIMA HAVE ASTERISK

Year	Jan	Feb	Mar	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1930	29·9	45·6	74 0	234	249	209	215	295	203	88·8	20·7	28·3
1931	40·5	74·9	77 0	200	215	*124*	262	210	*81·4*	95·5	31·7	15·7
1932	*17·0*	*22 5*	124	176	252	244	232	158	98 4	61·4	37·0	23·7
1933	49 1	50·3	141	255	280	296	263	176	108	—	—	—
1934	17·5	55·6	110	183	*120*	234	323	129	104	*33 3*	*14 5*	*14·5*
1935	24·6	46·6	*65·1*	166	178	141	183	*118*	142	36 3	37·3	20·1
1936	18 9	37 3	121	*144*	366	243	*135*	121	91 8	38·5	24 9	18·3
1937	21 9	28·4	87 3	148	231	300	232	189	135	43·7	18·6	24·9
1930	145	257	717	1230	1188	1206	1323*	1023	743	414	317	93 8
1931	160	327	545	928	1073*	931	832	701	556	306	165	87·0
1932	129	261	641	741	968*	966	718	599	409	323	135	86·7
1933	186	281	556	864*	784	703	768	687	409	—	—	—
1934	126	371	570	731	969*	909	795	765	608	305	179	81·1
1935	142	283	448	669	965*	931	784	670	543	383	108	89·4
1936	127	271	552	753	912	1002*	721	638	463	351	134	88·8
1937	105	240	544	663	941*	830	722	758	533	338	198	97·7
1930	82 1	141	315	682	719	725	781*	653	450	231	112	61·1
1931	94 9	163	317	476	625*	586	531	470	300	192	98·7	46·5
1932	69 8	135	339	505	557	648*	490	*380*	*260*	199	*69·3*	58·3
1933	91 0	154	350	520	568*	*527*	499	386	274	—	—	—
1934	60·7	176	333	497	730*	670	591	474	361	*173*	76·6	*45 6*
1935	64·8	133	*285*	469	*553*	577*	533	418	351	187	93·0	56·1
1936	68·4	139	286	460	639*	567	484	447	293	183	79 4	52·5
1937	*55·6*	*131*	319	*408*	568	589*	*466*	451	330	178	117	60·2

values during a month it becomes far harder to explain exceptionally high readings. It is even more difficult when dealing with measurements of the illumination integral, shown in kilolux-hours in Table III. The lowest values, 14·5 kl. hr occurred in November and December 1934, with cloudy skies darkened further by smoke. The greatest value, 1323 kl. hr , occurred on 7 July 1930, as noted previously. Apart from the high values of 1930, we find 1073 kl. hr. in May 1931, with nothing nearer than 969 in May 1934.

But it is the mean monthly illumination integrals that set 1930 apart as a most remarkable year, as may be seen from Table IV. From April 1930 to January 1931, every month save May and November is a maximum for the eight years, and as a rule is far ahead of the year ranking next. That November 1937 should exceed 1930, and December 1930 be only slightly greater than 1937 is further confirmation of the constancy of the cell. From Table V it may be seen that 1933 had far more sunshine than 1930. The air may have been rather cleaner in 1930 than in the other years since the rainfall was greater, but 1932 and 1935 were not far behind. The number of days with rain was also greatest in 1930, but one hesitates to advance a greater number of days with rain as an explanation of increased brightness

TABLE IV THE MAXIMUM VALUES OF THE MONTHLY VERTICAL ILLUMINATION INTEGRAL ARE SHOWN AS A PERCENTAGE OF THE CORRESPONDING MINIMUM VALUES. THE VALUES NEXT THE MAXIMUM ARE SHOWN SIMILARLY

	Year of maximum	Excess %	Year next maximum	Excess %	Year of minimum
Jan	1931	70	1933	63	1937
Feb.	1934	35	1931	24	1937
Mar.	1933	23	1932	19	1935
April	1930	67	1933	27	1937
May	1934	32	1930	30	1935
June	1930	38	1934	27	1933
July	1930	68	1934	27	1937
Aug.	1930	71	1934	24	1932
Sept.	1930	73	1934	39	1932
Oct.	1930	33	1932	15	1934
Nov.	1937	68	1930	61	1932
Dec.	1930	34	1937	32	1934

In the correlation of the exhaustion of the phosphate in sea water with sunshine in the spring during 1923, 1924 and 1925 owing to the multiplication of the planktonic algae (Atkins 1926, Table V and fig. 5) an indication was obtained of the need for more exact measurements of the illumination.

TABLE V. NINE MONTHLY AND ANNUAL TOTALS OF THE VERTICAL ILLUMINATION INTEGRAL, IN KILOLUX HOURS, WITH THE CORRESPONDING MEAN DAILY VALUE; MEAN DAILY SUNSHINE, IN HOURS; RAINFALL IN MILLIMETRES AND NUMBER OF DAYS WITH RAIN, AS RECORDED FOR PLYMOUTH HOE

	1930	1931	1932*	1933	1934	1935	1936*	1937
Kilolux hours, Jan.–Sept.	**138,670**	108,573	103,126	102,653	118,592	103,067	103,334	*101,143*
Kilolux hours, year	**151,100**	118,948	113,175	[113,000]†	127,666	113,353	113,013	*112,037*
Kilolux hours a day	**414**	326	309	310	350	311	309	*307*
Sunshine hours a day	4·38	*4·03*	4·28	5·17	4·88	4·51	4·27	4·25
Total rainfall	**1,115**	983	1,075	*782*	1,002	1,075	928	1,050
Days with rain	**209**	193	184	*169*	192	193	183	184
Ratio of daily kilolux hours to value for 1936, which is mean of 5 years	1·34	1·055	1·00	1·00	1·13	1·005	1·00	0·99
Ratio of daily hours of sunshine to value for 1936	1·03	0·945	1·00	1·21	1·14	1·055	1·00	0·995

* Leap year.

† This value has been taken as an approximation, since for 9 months 1933 and 1936 were much alike and the last three months of 1933 were sunny. This approximate total has been used to calculate the 1933 percentages shown in Table VI.

TABLE VI. VERTICAL ILLUMINATION INTEGRALS CALCULATED FOR EACH CALENDAR MONTH AS A PERCENTAGE OF THE TOTAL FOR THE YEAR

	1930	1931	1932	1933	1934	1935	1936	1937
Jan	1·68	2·47	1·91	2·50	1·47	1·77	1·88	1·54
Feb.	2·60	3·83	3·46	3·83	3·87	3·29	3·57	3·27
Mar.	6·47	8·25	9·27	9·60	8·09	7·80	7·83	8·85
April	13·54	12·00	13·38	13·81	11·70	12·42	12·22	10·92
May	14·76	16·28	15·23	15·59	17·72	15·12	17·54	15·71
June	14·40	14·80	17·17	13·98	15·74	15·26	15·05	15·78
July	16·03	13·82	13·40	13·68	14·33	14·57	13·30	12·90
Aug.	13·38	12·25	10·42	10·60	11·50	11·43	12·25	12·47
Sept	8·93	7·58	6·87	7·28	8·47	9·28	7·79	8·84
Oct.	4·74	5·02	5·45	—	4·20	5·07	5·02	4·93
Nov	2·22	2·49	1·84	—	1·80	2·46	2·11	3·12
Dec	1·25	1·21	1·60	—	1·11	1·53	1·44	1·67
Winter half	18·96	23·27	23·53	25·36	20·54	21·92	21·85	23·38
Summer half	81·04	76·73	76·47	74·64	79·46	78·08	78·15	76·62

See footnote, Table V.

In Table VI the illumination integrals are shown for each month as a percentage of the annual total. A correlation with changes in the plankton will be attempted elsewhere, but it is of interest to note that whereas the first three months of 1930 received 10·7 % of the annual illumination, in 1933 15·9 % was received, about 13–14·5 % being a more usual value. The summer half of the years received from three-fourths to four-fifths of the light. Contrary to what might be imagined, June is not always the brightest month, but was so in 1932, 1935 and 1937, with July at the top in 1930, and May in the other four years. The general absence of clouds and rain, with as a result less reflexion from the sky and a less transparent atmosphere, seems to be the cause of the low values for June.

Table VII shows the vertical radiation integrals for London. These were obtained with the Callendar radiometer, in which the radiation passes through glass, and are on the National Physical Laboratory scale of radiation. Guild (1937) has shown that this is almost identical with the most recent scale of the Bureau of Standards, Washington, and agrees with Abbot and Aldrich (1934) that the Smithsonian scale of 1913 gives values 2·3 % too high. As far as the complete years go, 1930 is at the top, though the excess over the minimum is only 19 %. The maximum radiation is usually in June, but in July for two years. Comparing the last line with Table IV, it may be seen that the excess percentage of the maximum is never quite as high as for Plymouth short wave daylight, and is markedly less in the brighter months One would expect that atmospheric impurities and particles serving as condensation nuclei would have a selective effect upon the shorter wave-lengths of sunlight, so that greater variations are to be expected in the region to which the sodium cell is sensitive. It is none the less difficult to see why in 1930 four of the brightest months, April, July, August and September, also January 1931 should be 67–73 % greater than the minimum, while June, October and December were maxima for the eight years also. Reference may also be made to the discussion in §§ IV and VI (Atkins and Poole 1936) and to Atkins, Ball and Poole (1937). In the 1936 paper the European pressure system was considered for April 1930, but no explanation was afforded for the exceptional values. It is possible that though the lower level winds were variable there may, during 1930, have been a predominance of arctic air of great purity, so that higher intensities of illumination were incident upon the lower strata of the atmosphere.

An entirely different explanation should not, however, remain unconsidered, namely, that the high values found relate only to the short-wave region, around 0·41 μ and that the sun's emission is especially variable

TABLE VII. VERTICAL RADIATION INTEGRAL AT LONDON, SOUTH KENSINGTON

Means for all available days, in gram calories per sq. cm. By courtesy of the Director of the Meteorological Office, Air Ministry. The last line shows the excess percentage of the maximum over the minimum as in Table IV

	Jan.	Feb.	Mar	April	May	June	July	Aug	Sept.	Oct.	Nov.	Dec.	Year
1930	37	*57*	155	176	260	337*	301	292	177	124	50	22	165
1931	37	71	147	186	262	324*	274	237	166	102	55	30	158
1932	—	—	152	193	227	335*	290	267	160	107¶	—	—	—
1933	—	—	—	—	—	—	—	—	201	106	42	*18*	—
1934	37	58	127	178	282	296§	339*‖	238	207	104	40	—	—
1935	—	—	129‡	194	280	297	326*	251	179	107	45	24	—
1936	29	67†	*112*	177†	250	*280**	*223*	*230*	*144*	94	*38*	29	139
1937	*28*	63	123	*149*	*224*	279*†	243	235	181	*79*	37††	24	139
Percent.	32	25	38	30	26	20	52	27	44	57	45	67	19

† 28 days only. ‡ 22 days only. § 27 days only. ‖ 25 days only. ¶ 24 days only.
†† 26 days only.

in this portion of the spectrum. Abbot has shown that there are periodical variations in the solar constant of radiation; the observations have recently been summarized and discussed by him (1935). Daily fluctuations are usually around 2 %, but may even show a range of 8 % (1920). Pettit (1932), on Mt Wilson, found large percentage changes in the ultra-violet around $0\cdot32\mu$, when compared with the green about $0\cdot5\mu$. An automatic recorder compared the effects of the two regions upon the same thermopile, the comparison being completed in 4 min., and immediately repeated. When computed for zero air mass the ratio of the intensities varied from 0·95 to 1·56. It is also claimed that these changes show positive correlation with the variations in total solar radiation. Pettit's work undoubtedly shows that one cannot measure short wave-lengths and assume that they always give a true measure of the variation near the middle of the spectrum, though it was shown by Atkins and Poole (1936, fig. 2 and discussion) that, for the regions covered by the cells used, this assumption was reasonably correct But the existence of notable exceptions lessens the value of the present series of records as far as their application to problems of photosynthesis is concerned. While the existence of such variations in the ultra-violet to green ratio has been established by Pettit, the validity of his computation to zero air mass has been questioned by Bernheimer (1933) on the basis of his own measurements of ultra-violet, extending from 1925 to 1933. He claims that these show an annual maximum in midwinter and a corresponding minimum in midsummer, variations due to the alterations in the turbidity of the atmosphere. He previously drew attention to the agreement between the yearly variation in the transmission of the atmosphere for $\lambda 3200$, as measured by Götz at Arosa, similar measurements at Upsala, the changes in the ultra-violet at Arosa for air-mass 2·9 and Pettit's measurements for the ultra-violet to green ratio From the similarity of the curves Bernheimer concluded that Pettit's deductions were untenable (1928, 1929, 1933 *b*).

I am indebted to Dr H. H. Poole for suggesting that a loose contact at the shunting resistance might lead to a fictitiously high current in the recorder and so explain the high values for 1930. These however occurred with each of the three separate shunt units, the connexions of which were soldered by the makers, no plugs being used. The same shunts are still in use.

I desire to express my indebtedness and tender my thanks to the following: The Government Grant Committee of the Royal Society for funds for the purchase of the Cambridge thread recorder and other instruments; to Dr H. H. Poole, with whom the first measurements were published, and to the Royal Dublin Society, in whose laboratories certain standardizations were carried out with him, to my colleagues and laboratory assistants at Plymouth who kindly helped in the daily changing of the charts, and otherwise, during the eight years, also to the Superintendents of the Meteorological Office, South Kensington, and of Kew Observatory, for the information they kindly supplied.

SUMMARY

A Burt vacuum sodium photoelectric cell was used with a Cambridge "thread recorder" to obtain daily records of the vertical illumination, around $0{\cdot}41\mu$. The greatest, least and mean values of the daily maximum are tabulated for each month, as are the corresponding measurements of the illumination integral in kilolux-hours. The constancy of the cell was established, and it was shown that the daily maxima varied from 3·2 to 197·6 kl on the carbon arc potassium cell scale, which for average daylight is close to the mean noon sunlight selenium cell scale.

The monthly mean of daily maxima was 150·9 for July 1930 and 12·4 for December 1931 From April 1930 to February 1931 the monthly means were, all save December, maximal for the eight years. The daily illumination integral varied from 14·5 to 1323 kilolux-hours, and the monthly means from 45·6 in December 1934 to 781 in July 1930, in this year the means for April, June, July, August, September, October, December, also January 1931 were maxima for the eight years Five years averaged 309 kilolux-hours a day, but 1930 showed 414, whereas 1933 had the highest sunshine average. No satisfactory explanation of the high values of 1930 is forthcoming, but November 1937 exceeded the high November 1930 The vertical radiation integral, as measured at London, showed that 1930 was rather higher than other years Possibly the explanation lies in the persistence of exceptionally clear upper air during 1930, or perhaps in an actual increase in the ratio of ultra-violet to green in the solar spectrum as found by Pettit, a result contested by Bernheimer. The six summer months receive from three-quarters to four-fifths of the annual daylight, of this the brightest month (May, June or July) receives 15–18% and December 1·5%.

REFERENCES

Abbot, C. G. 1920 *Proc Nat. Acad. Sci., Wash.*, **6**, 674–78.
— 1935 *Smithsonian Misc. Coll.* **94**, No. 10.

Abbot, C. G. and Aldrich, L. B. 1934 *Smithsonian Misc. Coll.* **92**, No. 13.
Atkins, W. R. G. 1926 *J. Mar. Biol Ass. U.K.* **14**, 447–67.
Atkins, W. R. G., Ball, N. G. and Poole, H. H 1937 *Proc. Roy. Soc.* A, **160**, 526.
Atkins, W. R. G. and Poole, H. H. 1936 *Philos. Trans.* A, **235**, 245–72.
Bernheimer, W. E. 1928 *Naturwissenschaften*, **16**, 26–7.
— 1929 *Handbuch Astrophys.* **4**, 45–48. Berlin. J. Springer.
— 1933 *a* *S.B. Akad. Wiss. Wien*, **142**, 449–56.
1933 *b* *Handbuch Astrophys.* **1**, 460–3
Guild, J. 1937 *Proc. Roy. Soc.* A, **161**, 1–37.
Pettit, E. 1932 *Astrophys. J.* **75**, 185–221
Poole, H. H. and Atkins, W. R. G. 1935 *Philos Trans.* A, **235**, 1–27

Quantitative spectrographic analysis of biological material

III. A method for the determination of sodium and potassium in glandular secretions

By J. S. Foster, F.R S., G. O. Langstroth, Ph.D.
and D R. McRae, Ph.D

Department of Physics, McGill University, Montreal, Canada

(*Received* 15 *December* 1937)

A study of certain digestive glands in experimental animals (Langstroth, McRae and Stavraky 1938) required the analysis of samples of secretion, ranging from less than 1 to 15 c c in volume, for several different substances. The use of the quantitative spectrographic method of analysis for Na and K described in this article aided considerably in obtaining the data for each sample, and so resulted in a description which was relatively complete as compared with that obtained under a purely chemical procedure. This more complete knowledge of the composition of the secretions was found to be of the utmost importance in the interpretation of the results, and in obtaining an understanding of the behaviour of the glands.

The method of analysis for Na and K, like that for Pb described in paper I (Foster, Langstroth and McRae 1935), depends on the determination of the intensity ratio of a chosen line of the investigated element to one of an internal standard element. In contrast to the method of paper I, however, the sample is excited in a condensed A.C. spark discharge, and use is made of a standard working curve, viz. the intensity ratio plotted against concentration. Some important features of the procedure are as follows. (*a*) The load placed on the electrode contains a fixed "large" amount of some suitable added salt (the buffer). This serves to obviate variations in the operation of the source due to variations in the composition of the samples It permits the use of one working curve in the analysis of samples which vary considerably in composition. (*b*) The sample may be placed directly on an electrode previously prepared by drying on it a solution containing the buffer salt and internal standard. With this procedure only 0·01 c.c. of a sample is required for an analysis of both elements. (*c*) The spectrum of each sample is photographed with an antimony absorption step-weakener (Langstroth and McRae 1937) before the spectrograph slit. The photographic blackenings of investigated lines then lie in the normal exposure region in some steps, and can there be accurately measured. In this way one exposure is sufficient for the determination of both elements, or of several elements if required.

The method has been applied in determining Na concentrations between 1×10^{-4} and 30×10^{-4} g./c.c , and K concentrations between 1×10^{-4} and 12×10^{-4} g./c.c. The probable error of the mean of a pair of determinations, which requires only 0·02 c.c. of a sample, is about 4 % The method for K gives essentially the same results as Kramer and Tisdall's chemical method (1921), but there is evidence to show that it is more reliable, at least under our test conditions. The low risk of contamination or loss, and the lack of dependence of determinations on the chemical form in which the element is present or on the extraneous composition of the sample, are important characteristics of the method. As pointed out in the introductory paragraph, the small size of the sample required for an analysis is a decided advantage in certain problems. A pair of determinations for Na and K may be made in three-quarters of an hour.

1. PROCEDURE

The sparking circuit used for exciting the samples is shown in fig. 1. The inductance L is about 2000 μH, and the condenser C has a capacity of about 0·02 μF. The transformer supplying the power has a 1 : 200 ratio, and is operated with 15 V across the primary. Examination of the discharge by means of a rotating mirror shows that with this arrangement sparks usually pass between the electrodes only on every other half-cycle of the 60 cycle A.C. supply.

The electrodes consist of a copper point and a copper plane (9 × 10 mm.) on which the load is placed. These are mounted in a holder which permits the plane to be so moved during an exposure that the spark passes over the entire surface, as described in paper I. The electrode separation employed is 3·6 mm. The image of the source magnified five times is projected on the slit of the spectrograph by a condensing lens; accordingly only light coming from a definite region of the discharge midway between the electrodes is examined.

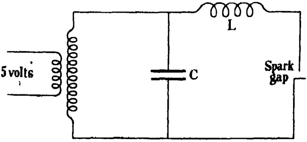


FIG. 1 Sparking circuit.

An absorption step-weakener of antimony (Langstroth and McRae 1937) is placed immediately before the slit of the large quartz spectrograph (Foster 1936) used in this work. One step of the weakener reduces the intensity of the chosen Na line λ 3302·3 so that the photographic blackening produced by it is comparable to that produced by the K line λ 4044·7 as transmitted by another step, both lying in the normal exposure region.

The copper electrodes are dipped in concentrated nitric acid and rinsed with distilled water. Glassware is cleaned by immersing it in boiling nitric acid, and washing in distilled water

In making a determination, 0 03 c.c. of the buffer solution is placed on the plane electrode by means of a specially made pipette, and is dried in a dessicator. The buffer solution contains the internal standard elements as well as the buffer salt. The buffer salt used in our work is lithium tartrate at a concentration of 0·033 g./c c. Pb and Cd are used as internal standards for K and Na respectively; their concentrations are such that Pb λ 4057·8 and K λ 4044·7, and Cd λ 3261·0 and Na λ 3302·3 produce comparable blackenings on the photographic plate. A definite amount (0·01 c.c.) of the sample is placed in a uniform layer over the surface of the prepared electrode by means of a pipette which will be referred to later, and this new material is dried as before.

The electrodes are sparked and the spectrum photographed. In making an exposure the plane electrode is so moved that the spark passes four times over its surface at a constant rate each time, in 60, 30, 15 and 15 sec. respectively. Calibration marks are put on the plate in the λ 3302 and λ 4044 regions with a step-slit and quartz band lamp.

The intensity ratios λ 4044, K : λ 4057, Pb, and λ 3302, Na · λ 3261, Cd, are determined from the plate by the standard microphotometric method. The following technical procedure has, however, been adopted. A Moll microphotometer, used as a direct reading instrument, has been so modified that the light from the galvanometer falls directly on semi-logarithmic graph paper, and that the clear plate deflexions may be conveniently adjusted to read 100 on the linear scale With this arrangement, no numerical values for galvanometer deflexions are read, those for the calibration marks are marked at the appropriate intensities on the logarithmic scale, to give the calibration curve for the plate; those for the spectral lines may be read off directly in terms of intensity from this curve The procedure results in a considerable saving of time and labour.

The concentrations corresponding to the determined intensity ratios are read off from standard working curves. These curves are constructed by plotting determined intensity ratios for known solutions against the known concentrations.

2 THE PRECISION AND ACCURACY OF THE DETERMINATIONS

The precision of the method as used in analyses of saliva and pancreatic secretion is illustrated in Table I. Two determinations for each element were made on each sample, and the average was taken as the representative value The deviations of Table I refer to the deviation of either determination from this value

TABLE I

Method of mixing	No of samples	Mean deviation (%)		Standard deviation (%)		Probable error (%)	
		Na	K	Na	K	Na	K
(*a*) In test-tube	116	4·0	5·1	4·8	6·2	3·2	4·2
(*b*) On electrode	19	3·8	5·4	4·8	6·9	3·2	4·7

When the sample is placed on a prepared electrode as described above, the deviations include the errors in measurement of the volumes of sample and buffer solution used. On the other hand, earlier analyses were made by mixing 0·25 c.c. of the sample with an appropriate volume of the buffer solution in a test-tube, and using this mixture for both determinations. Under these circumstances, the deviations exclude the errors in measurement of the volumes of sample and buffer solution.

It is apparent that the errors introduced by errors in the volume measurements in procedure (*b*) are negligible in comparison with the remaining errors inherent in the spectroscopic method. Procedure (*b*) is more rapid than (*a*), and requires a much smaller volume of the sample. In view of the results of Table I the precision of the method is represented by a probable error of 3·2 % for Na, and 4·7 % for K.

Table II makes a comparison between K concentrations in saliva as determined spectroscopically, and as determined by Kramer and Tisdall's (1921) chemical method.

TABLE II

K concentration (mg. %)		
Spectroscopically	Chemically	Deviation from mean (%)
26	21	10·6
28	28	0·0
30	26	7·2
29	32	4·9
34	38	5·6
44	43	1·1
40	42	2·4
45	42	3·4
47	50	3·1
Sum 323	322	

The fact that the sums of the two sets of determinations differ by only 0·3 % indicates that no constant error is present in either method, or that a similar constant error is present in both. In view of the totally different character of the two methods the latter alternative appears to be highly improbable.

The deviation of individual chemical and spectroscopic determinations from their mean is in one case as high as 10·6 %. Such large deviations probably result from large errors in the chemical determinations, since, as indicated by the following illustration, the spectroscopic method is the more reliable under our test conditions. Each of several samples of saliva was divided in half, and to one half was added a known amount of a K salt. All half-samples were analysed chemically and spectroscopically for K, and the added amount, as determined from the analyses, was compared with the known added amount. It was found that the average deviation from the known value was 17 % for the values determined chemically, while for those determined spectroscopically it was only 5 %. Similar tests were made spectroscopically for Na, with similar results.

3. DISCUSSION

The problems of quantitative spectrographic analysis may be considered to fall into three distinct classifications: (a) problems concerned with the preparation of the sample for excitation, (b) problems concerned with the excitation of the sample, and (c) problems concerned with the measurement of relative intensities in the emitted spectrum.

The preparation of the sample. It is desirable that the preparation of the sample should involve as little treatment as possible, in order to minimize the chance of contamination or loss of material. This condition appears to be satisfactorily fulfilled by the present procedure, which involves only the transfer of a small quantity of the sample from the container in which it is collected to the electrode surface.

It is necessary that the volume delivered by the pipette to the electrode be closely reproducible. Pipettes with long fine points slightly curved at the end are used for this purpose. The curve facilitates the removal of liquid which tends to adhere to the outside of the pipette after delivery of the sample. As indicated by the figures of Table I, 0·01 c c. of liquid may be placed on an electrode with a high degree of reproducibility with these pipettes.

The excitation of the sample. If reliable results are to be obtained, it is necessary that the intensity ratio of a line of the investigated element to some internal standard line be dependent only on the amounts of the two elements present, and on no other factor which may vary from sample to sample. The degree to which this condition is fulfilled determines the success of a spectroscopic method.

If a sample is placed alone on an electrode, the intensity ratio obtained for lines coming from different elements will depend on the amount of extraneous matter present. For example, the intensity ratio of the Pb and K lines used in this work increased by 50 % when the deposit of lithium tartrate on the electrode was changed from 0·3 to 1·0 mg./cm.^{2} of surface; the change in the Na : Cd ratio was somewhat less. In our experiments, the samples of glandular secretions to be analysed contained in some cases as much as 3 %, and in others as little as 0·3 % solids. Because of this variation in total solids, accurate determinations could not be made for such samples by placing them alone on the electrode. To overcome the difficulty, some suitable substance is deposited on the electrode in a fixed amount greatly in excess of the amount of solids in the samples. The discharge then takes its character from this added substance (the buffer), and is little affected by variations in the composition of the samples. Under these conditions one may use the same working curve for different samples. It is obvious that in determining the working curves, the deposit of buffer salt on the electrode must be the same as that used in actual analyses. Lithium tartrate has been found to be a satisfactory buffer salt for our work. It dries on the electrode in a closely adhering layer, which is "burned" rather than flaked off as is the case with many inorganic salts.

In view of the preceding statements, it is clear that the deposit on the electrode must form a layer of nearly uniform thickness if reliable results are to be obtained By drying rapidly at low pressure some samples of saliva may be made to form a thick deposit at the edges of the electrode, leaving the centre nearly bare. The K determination obtained from such a deposit may differ by as much as 40 % from the determination made from a uniform deposit Thus far no difficulty has been encountered in obtaining sufficiently uniform deposits when the sample is dried in a desiccator.

The condenser in the sparking circuit is large enough to produce a spark which will burn through the deposit on the electrode. The voltage across the primary of the transformer is reduced to a point slightly above that at which a discharge just takes place between the electrodes. Under these conditions the general heating of the electrodes is a practical minimum, and the destruction of the character of the deposit in places not immediately attacked by the spark appears to be negligible. The adjustment of the electrode spacing is not critical; a 25 % change in separation causes less than 7 % change in the intensity ratios.

There is evidence to indicate that the excitation in the discharge column of the source used in this work is thermal in character (Langstroth and McRae 1938), i.e. the average populations of the initial states are in a Boltzman distribution. It is therefore desirable to choose internal standard lines which originate in initial levels having nearly the same excitation potentials as those of the investigated lines of the elements to be determined. Energy level data for the lines used in our experiments are given in Table III.

Consideration shows that if the working curve was determined with a discharge temperature of 9000° K., and an observation was made with the discharge at 8000° K., the error in the determination of K due to this variation would be 20 %. The error in the Na determination would be considerably less, since the Na and Cd excitation potentials differ by only 0·07 e-volts whereas the corresponding difference for Pb and K is 1·32 e-volts. The fact that the average error is 1 % greater in the K, than in the Na determinations (Table I), may be due in part to fluctuations in the discharge temperature of the source. An average fluctuation of 50° from the standard temperature would account for a 1 % greater deviation in the K measurements. Another factor contributing to this greater deviation for K is referred to below.

TABLE III

		State		Excitation potential e-volts
Element	Line	Initial	Final	
Na	3302·3	(3p) $^2P_{\frac{3}{2}}$	(1s) $^2S_{\frac{1}{2}}$	3·71
K	4044·7	(3p) $^2P_{\frac{3}{2}}$	(1s) $^2S_{\frac{1}{2}}$	3·04
Cd	3261·0	(2p) $^3P_1$	(1s) $^1S_0$	3·78
Pb	4057·8	[6p($^2P_{\frac{1}{2}}$) 7s] 2^0_1	(6p^2) $^3P_2$	4·36

Measurement of intensity. In our applications of the method, the background intensity was considered to be sufficiently small to be neglected. It appeared to be somewhat greater however in the region of the K line than in that of the Na line. This circumstance, combined with the fact that the blackening for the K line was small as compared with that of the other lines, results in a greater error in the determination of the K : Pb ratio. Hence a greater mean deviation in the K determinations is to be expected. This is found, as shown in Table I. The fact that the K determinations of the second row of Table 1 have a greater probable error than those of the first row (4·7 % as compared to 4·2 %) is due to the fact that for these determinations the exposures (and so the background) were twice as great, while the blackenings of the K line remained about the same. It may be noted that the probable errors for Na, for which the background is a much smaller consideration, are the same in the two rows of Table I.

We are indebted to Dr G. W. Stavraky for making the chemical analyses referred to in this article. We are also indebted to the Rockefeller Foundation for financial assistance.

SUMMARY

An internal standard method of quantitative spectrographic analysis, as applied in the determination of Na and K in glandular secretions, is described. The probable error in a determination is about 4 %.

By adding an appropriate foreign substance to the sample in such quantities that the character of the condensed A.C. spark discharge is determined by this substance, it is possible to use the same working curve for samples of considerably different composition.

The sample and internal standard may be separately placed on the electrode by a pipette specially designed for accurate delivery of small volumes. The volume of a sample required for the determination of both elements is thus reduced to 0·01 c.c.

By photographing the spectrum of the sample with an antimony absorption step-weakener before the spectrograph slit it is possible to obtain determinations for two or more elements from a single exposure

REFERENCES

Foster, J. S. 1936 *Canad. J. Res.* A, **14**, 173–6.

Foster, J. S., Langstroth, G. O and McRae, D. R. 1935 *Proc. Roy. Soc.* A, **153**, 141–52.

Langstroth, G. O. and McRae, D. R. 1937 *Canad. J. Res.* A, **15**, 154–60.

— — 1938 *Canad. J. Res.* A, **16**, 17–27.

Langstroth, G. O., McRae, D. R. and Stavraky, G W. 1938 *Arch. int Pharmacodyn.* **58**, 61–77.

Kramer, B. and Tisdall, F. F. 1921 *J. Biol Chem* **46**, 339–49.

Thermal conduction in hydrogen-deuterium mixtures

BY C. T. ARCHER, A.R.C.S., M SC.

Senior Lecturer in Physics, Imperial College of Science, London, S.W.7

(*Communicated by A. O. Rankine, F R.S —Received* 30 *November* 1937)

INTRODUCTION

The work to be described here was undertaken by the author in order to provide accurate data on thermal conduction in gaseous mixtures of deuterium and hydrogen at ordinary temperatures, to be applied to the analysis of such mixtures by the thermal conductivity method developed by A. Farkas and L. Farkas (1934) and others; and also to make an accurate direct determination of the thermal conductivity of pure deuterium.

With regard to the latter quantity, the values obtained by other workers to date vary widely, showing a maximum difference of the order of 11 %.

Van Cleave and Maass (1935), using a relative hot-wire method and assuming the value of the thermal conductivity of pure hydrogen at 0° C. (K_0) to be 0·000416 cal. cm.$^{-1}$ sec $^{-1}$ deg.$^{-1}$ C., gave as their final result for deuterium

$$K_0 = 0{\cdot}000295 \pm 0{\cdot}000003 \text{ cal. cm.}^{-1} \text{sec.}^{-1} \text{deg}^{-1} \text{C.}$$

A later determination made by Kannuluik (1936), using a modified hot-wire method, gave the value

$$K_0 = 0{\cdot}000329_4 \text{ cal cm.}^{-1} \text{sec.}^{-1} \text{deg.}^{-1} \text{C.}$$

Again, an extensive series of determinations of the thermal conductivities of various gases, including deuterium, has been made by Nothdurft (1937), also using a hot-wire method, and in this case the value obtained was

$$K_0 = 0{\cdot}0003031 \pm 0{\cdot}0000011 \text{ cal. cm.}^{-1} \text{sec}^{-1} \text{deg.}^{-1} \text{C.}$$

A preliminary announcement of the value obtained by the author (Archer 1936) has already been made, and it is pointed out that the two former values quoted above were published during the course of the work to be described here, whilst the last quoted was announced after the thermal conductivity determinations had been completed. Unfortunately, however, circumstances beyond the control of the author seriously hindered the work on the determinations of the concentrations of deuterium oxide and water, used in the preparations of the gases, and in consequence a delay in the final publication of the results has occurred.

DESCRIPTION OF APPARATUS

The apparatus used is shown diagrammatically (fig. 1) and will be described briefly.

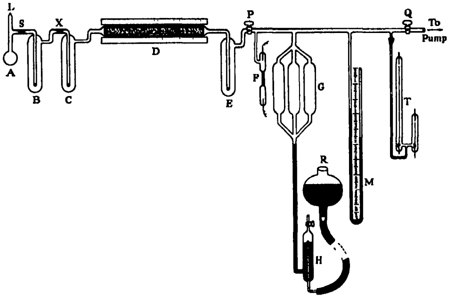


FIG. 1

A is a small bulb of pyrex glass communicating with the pyrex glass liquid-air traps, *B* and *C*. *C* is connected, by using joints of intermediate glasses, to a quartz tube placed inside an electrically heated cylindrical furnace, *D*, and containing pure magnesium turnings. This tube, again by using joints of intermediate glasses, is connected to a third pyrex glass liquid-air trap, *E*, and thence through the tap, *P*, to a discharge tube, *F*, a compression apparatus, *G*, a mercury manometer, *M*, the thermal conductivity tubes, *T*, and finally through the tap, *Q*, to the vacuum pumps used.

The compression apparatus consists of three pyrex glass cylindrical bulbs, *G*, connected as shown, together with a mercury reservoir, *R*, and an air trap, *H*, the latter to prevent air from being carried into the apparatus by the mercury when the reservoir, *R*, is operated. The function of the apparatus was to overcome the difficulty of being able to prepare only a small volume, about 150 c.c., of the various gases at a reduced pressure. The gas having been prepared at a fairly low pressure, about 100 mm. of mercury, with the reservoir, R, in its lowest position, could be compressed by gradually raising the reservoir to fill the bulbs, G, with mercury and the pressure thereby increased to about 700 mm. of mercury.

The magnesium turnings, after having been thoroughly washed in ether to remove all traces of grease, were contained in a thin walled quartz tube, open at both ends and lined with asbestos, placed inside the slightly larger quartz furnace tube. This arrangement was found to be necessary because when a single furnace tube was used a thin layer of magnesium was deposited on the inner walls of the furnace tube after heating and caused the tube to be fractured.

The furnace could be raised to a temperature of about 650° C., the temperature being indicated by means of a copper-constantan thermocouple placed between the quartz tube and the furnace wall. It was found possible to maintain the temperature of the furnace constant to within 5° C. of the desired temperature, usually 500° C., for several hours by controlling the electric current with suitable rheostats

The thermal conductivity tubes, T, were of the type employed by the author in previous work of a similar nature, the tubes being made of ordinary glass and connected to the rest of the apparatus by a carefully ground joint. The main tube and compensating tube were cut from the same piece of selected and calibrated tubing, the internal radius (r_2) being 0·6646 cm. and the external radius (r_3) 0·7692 cm. The thin wires sealed in the tubes were of pure platinum, the radius (r_1) being 0·003977 cm.

The gas pump system consisted of a motor driven Hyvac oil pump as backing pump and a gas heated mercury vapour pump.

PREPARATION OF GASES

In the greater part of the work, the apparatus described above was used to prepare the various mixtures of deuterium and hydrogen, as well as pure hydrogen and also deuterium as pure as it was possible to obtain. The first preparation was that of hydrogen as a check on the method, specially prepared distilled water being used for the purpose, and the method of preparation will now be described.

First, the whole apparatus was evacuated, the pumps being kept in operation for several hours while the electric furnace was heated to about 600° C., the discharge tube, F, being used to indicate the vacuum condition. This prolonged treatment and the heating of the magnesium to a temperature higher than that maintained in the actual preparation of the gases were found to be necessary in order to degas completely the apparatus and the magnesium. Moreover, it was found necessary to repeat the prolonged degassing process after each preparation; traces of hydrogen were observed to be present during the first few hours of the heating, and the process was continued until the discharge tube indicated the complete absence of all gases. Without this precaution it could not be ensured that the gas obtained was of the same concentration as in the liquid used for the preparation.

After allowing the furnace to cool to atmospheric temperature, the tap, *P*, was closed, the seal at *L* broken, and about 0·5 c.c. of the liquid was introduced into the bulb, *A*, and the seal at *L* was remade. The bulb, *A*, was surrounded with liquid air in order to freeze the liquid, the tap, *P*, opened and the apparatus again evacuated The trap, *B*, was next surrounded with liquid air, the frozen specimen in *A* allowed to melt and evaporate slowly, being again frozen in *B*. On the completion of this distillation the bulb, *A*, was removed at the constriction, *S*. The specimen was then distilled to the trap, *C*, by a similar method, and the trap, *B*, removed at the constriction, *X*. In this manner any gases dissolved in the liquid were removed and any traces of solid impurity left behind in either *A* or *B*.

Keeping the specimen frozen in *C*, the furnace was next heated until a steady temperature of about 500° C was attained, and at this stage the specimen in *C* was allowed to melt and evaporate very slowly, the pump system now being cut off by closing the tap, *Q*. Thus the vapour passed over the heated magnesium and was decomposed, the required gas passing over into the whole apparatus. Any slight traces of liquid not decomposed were collected by freezing in the trap, *E*, which was immersed in liquid air throughout the preparation. By careful control of the evaporation, however, from *C* it was found possible to decompose the whole of the specimen, some eight hours usually being occupied in the process. During this time the mercury reservoir, *R*, was kept in its lowest position so that the bulbs, *G*, and the tubes, *T*, were filled with the gas at a final pressure of about 100 mm of mercury, the pressure being indicated by the manometer, *M*. At this stage the tap, *P*, was closed.

Exactly the same method was employed in preparing the various mixtures of deuterium and hydrogen from mixtures of deuterium oxide and distilled water, using deuterium oxide of 99·95 % guaranteed concentration, supplied by Messrs Imperial Chemical Industries, Ltd. The liquid mixtures were made of an approximate concentration only of deuterium oxide in water, the exact concentrations being determined at a later stage.

Lastly, deuterium was prepared by the same method from a sample of the 99·95 % deuterium oxide.

In every case the preparation was repeated, using fresh samples of the various liquids, and the thermal conductivity observations on each gas specimen were repeated until in the case of each gas specimen consistency was attained. For this purpose the current measurements (see below) were used, an agreement to within 0·0005 amp. in the values being considered sufficiently consistent, the values ranging from 0·46 to 0·27 amp. in the whole series of observations.

As a final check on the thermal conductivity observations for deuterium, a sufficient supply of the gas was prepared by the decomposition of a specimen of the 99·95 % deuterium oxide by metallic sodium, using a similar form of apparatus and the method described by Mann and Newell (1937).

THERMAL CONDUCTIVITY OBSERVATIONS

The form of thermal conductivity apparatus used in the present instance is the vertical compensated hot-wire system evolved as the result of an extensive series of experiments on thermal conduction in gases, using the hot-wire method, in which the author has taken part, the main object of the system being to eliminate as far as possible the effect of losses of heat by convection.

In the first series of experiments (Gregory and Archer 1926*a*), two compensated hot-wire systems were used, the tubes being of different radii and placed horizontally. In this case there were convective losses in both systems, being much greater in the wider tubes than in the narrower. It was found possible, however, to eliminate the effects of convection by observing the pressures in both systems at which such losses vanished, the temperature conditions in the two systems being identical.

It had been shown previously (Weber 1917) that in a similar hot-wire system with the tubes placed vertically the heat losses by convection were very much smaller than in the same system placed horizontally. Hence it was decided to work with vertical tubes, modified in such a way as to reduce the convective losses still further. Such a modified hot-wire system was first used (Gregory and Archer 1926*b*) to show that the thermal conductivity of air is independent of pressure, the results indicating that convection was almost entirely eliminated over the range of pressures and temperatures used in that instance. The double system of the modified vertical type, however, was adopted for further experiments, being used to determine the thermal conductivity of carbon dioxide (Gregory and Marshall 1927), of oxygen and nitrogen (Gregory and Marshall 1928), and of carbon monoxide and nitrous oxide (Gregory and Archer 1928). Then, a single vertical system was first used for direct determination in an investigation of the thermal conductivities of the saturated hydrocarbons in the gaseous state (Mann and Dickins 1931).

Up to this time no account had been taken of the effect of accommodation, and the consequent discontinuity of temperature between a gas and a solid surface, on thermal conduction in gases in the work detailed. It was realized, however, that the effect though small should be taken into account, and consequently with this object in view a re-examination of the results of the work first mentioned above was made (Gregory and Archer 1933). It was then established that by using a relation of the type adopted in the present work (see below), the two effects of convection and temperature drop could be successfully eliminated from the thermal conduction, a single vertical hot-wire system being sufficient for the purpose. This was further confirmed in the case of a series of gases (Dickins 1934) of which the thermal conductivities and the accommodation coefficients relative to a platinum surface were determined The method, with some modification in the theoretical treatment, has since been applied in the case of hydrogen over a range of temperature up to 300° C. (Gregory 1935), and also in the case of carbon dioxide over a similar range of temperature (Archer 1935)

The construction and method of use of the thermal conductivity apparatus need little further description here, the calibration, etc , of the hot-wire tube system following exactly the same lines as described in the papers on the work mentioned above.

The tube system was maintained at a constant temperature, 0° C., throughout the whole series of observations, by means of the usual form of ice-bath fitted with a motor-driven stirrer, the platinum wires being connected to a Callendar-Griffiths bridge, used in conjunction with a Tinsley thermoelectric potentiometer.

The gas specimen having been prepared, the reservoir, *R* (fig. 1), was raised to obtain the maximum possible pressure in the tube system, usually about 700 mm. of mercury. The Callendar-Griffiths bridge having been set to balance for a resistance corresponding to a predetermined temperature of the platinum wires in the tube system, the strength of current passing through the wires was adjusted by means of sensitive rheostats until the bridge was balanced. The strength of the current was then measured by means of the potentiometer, and at the same time the pressure of the gas in the apparatus was observed by using the mercury manometer, *M*.

Next, the pressure of the gas was reduced by lowering the mercury reservoir, *R*, the strength of the current adjusted to restore balance of the bridge and the observations repeated.

This procedure was carried out at a series of pressures of the gas, the mercury reservoir, R, being lowered in suitable stages until the minimum possible pressure, usually about 100 mm. of mercury, was attained, the strength of current being measured and the pressure observed at each stage.

Three other similar sets of observations were made, with the Callendar-Griffiths bridge setting adjusted to correspond to three other temperatures of the platinum wires in the tube system, using the same specimen of gas over the same range of pressures in each set of observations.

This procedure was repeated in the case of each gas specimen prepared, four sets of observations being made for each gas with the heated platinum wires in the tube system at four different temperatures ranging from 21·644 to 10·370° C.

CALCULATIONS

In the present work, the relation

$$\frac{1}{Q} = \frac{\log_e r_2/r_1}{2\pi KJl\theta} + \frac{A}{P.\theta} \tag{1}$$

was used to determine the thermal conductivity, K, of the gas.

In this relation, Q represents the loss of heat per second by conduction, in the absence of convection, from an electrically heated platinum wire of radius r_1, of effective length l, mounted coaxially with a glass tube of radius r_2, and compensated for "end" effects, etc., θ being the difference of temperature between the platinum wire and the inner wall of the glass tube. P is the pressure of the gas in the tube system, and A is a constant involving the accommodation coefficient of the gas.

The values of Q were calculated from the observations of the effective resistance of the platinum wire and the strength of the current required to maintain the temperature of the wire constant during the observations.

The values of θ were also obtained from the observations of the effective resistance of the platinum wire, the necessary corrections to the platinum scale temperatures being applied to obtain the corresponding Centigrade temperatures; allowance also was made for the flow of heat through the walls of the glass tubes in each case, corrections being calculated from the internal (r_2) and external (r_3) radii of the tubes and the thermal conductivity of the material.

By plotting the values of $1/Q$ against those of $1/P$, a straight line is obtained. The intercept of this line on the $1/Q$ axis represents the quantity

$$\frac{\log_e r_2/r_1}{2\pi KJl\theta},$$

from which the value of the thermal conductivity, K, of the gas can be calculated.

The observations obtained for each of the four settings of the Callendar-Griffiths bridge were treated thus, the value of K found in each case being that at the average temperature between that of the heated wire and that of the internal wall of the glass tubes. The value of the thermal conductivity at 0° C., K_0, was then obtained by extrapolation from the four results.

At the same time, the slope of each straight line gives the value of the quantity, A/θ, corresponding to the particular experimental conditions, and by using the appropriate expression for A the accommodation coefficient of the gas at the temperature of the heated platinum wire can be obtained. It has been shown recently by Gregory (1936) that in the general equation (1) above, A may be expressed in the form

$$\frac{1}{2\pi l}\frac{\sqrt{(2\pi M)}}{R}\frac{2-\alpha}{2\alpha(\beta+\frac{1}{2})}\left\{\frac{\sqrt{T_1}}{r_1}+\frac{\sqrt{T_2}}{r_2}\right\},$$

in which l is the effective length of the heated platinum wire, M the molecular weight of the gas, R the gas constant, α the accommodation coefficient and β the specific heat per molecule of the gas, T_1 and r_1 the absolute temperature and radius of the platinum wire, T_2 and r_2 the absolute temperature and radius of the inner wall of the glass tubes.

This expression was used to determine the accommodation coefficient in the cases of deuterium and hydrogen. Experimental values of the specific heat per molecule for hydrogen were available, but in the case of deuterium no such data are available, and the classical value, 2·5, was used in the calculations.

In the present instance, calculations of the heat lost by radiation from the heated platinum wire in the prevailing experimental conditions showed that such losses were negligible in comparison with the total heat losses, and in view of the estimated accuracy of the observations the corrections were not applied.

OBSERVATIONS AND RESULTS

Space will not permit the reproduction of all the necessary observations taken in the course of the work, nor is it considered desirable to quote in detail all the experimental data. As an illustration, however, a typical set of lines are reproduced graphically in fig. 2. These particular lines were obtained from the experimental observations made in the case of deuterium.

These lines are typical of the results generally and show clearly the adherence of the observations to the straight line law according to relation (1) above.

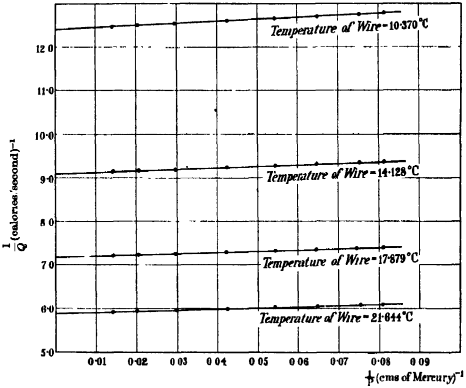


FIG. 2

The final value of the thermal conductivity of each of the various samples of gas, calculated from the whole series of experimental observations, is contained in the table below:

Percentage Hydrogen	 Deuterium	K_0	Temperature coefficient
100	0	$0·000418_3$	$0\ 0026_1$
80 2	19 8	$0\ 000382_4$	
65 5	34 5	$0·000364_7$	
49 6	50 4	$0\ 000350_4$	
39 5	60·5	$0·000341_1$	
18·7	81·3	$0·000323_1$	
0 05	99 95	$0·000308_3$ (*a*)	$0\ 0030_1$
		$0·000307_8$ (*b*)	$0·0029_7$

In this table the figures in the first two columns show the percentage concentrations of hydrogen and deuterium present in the different gas specimens used. These concentrations were obtained from a series of determinations of the densities of the liquid mixtures used in the preparation of the gases. For this purpose a flotation method, devised by Dr R. H. Purcell of the Chemistry Department of the Royal College of Science, was used, a composite float of silica and glass of about 0·25 c.c. capacity being filled with the liquid and the flotation temperature in a bath of pure degassed distilled water observed.

In the third column are shown the values of the thermal conductivity of each gas specimen, in each case at the temperature of 0° C. and expressed in the usual units, cal. cm $^{-1}$ sec.$^{-1}$deg.$^{-1}$ C. In the case of deuterium, the value marked (*a*) is that obtained from the gas prepared by the magnesium method, while that marked (*b*) is that obtained from the gas prepared by the sodium method.

The fourth column contains the values of the coefficient of increase of thermal conductivity with temperature, between 0° C. and an average mean temperature of the gas of 10·9° C., in the case of hydrogen and deuterium.

The values of the thermal conductivity are also represented graphically in fig. 3, the values of K_0 from the above table being plotted against the percentage concentrations of the gases in each case.

The accuracy of the thermal conductivity results is estimated to be of the order of 0·25 %.

Finally, the value of the accommodation coefficient also was calculated in the case of pure hydrogen and of deuterium, the results obtained being:

Hydrogen	...	$0{\cdot}29_6$,
Deuterium	...	$0{\cdot}37_6$.

These values are relative to a platinum wire surface at a temperature of 0° C., and are in good agreement with similar results obtained previously by other observers.

In conclusion, the author wishes to express his sincere thanks to Professor Thomson for his encouragement and the facilities to carry out the work, to Professor Rankine for his enthusiastic interest and helpful advice, and to Dr R. H. Purcell for the valuable help afforded in the preparation of the gases and in the determination of the concentrations.

SUMMARY

The paper describes an experimental investigation of thermal conduction in hydrogen, deuterium, and hydrogen-deuterium mixtures of varying concentration. The hot-wire method, used by the author and others

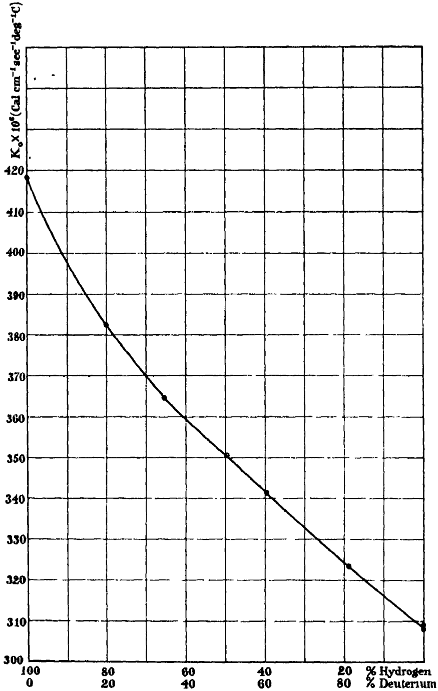

$K_o \times 10^6$ (Cal cm^{-1} sec^{-1} deg^{-1}C)
420
410
400
390
380
370
360
350
340
330
320
310
300
100
0
80
20
60
40
40
60
20
80
% Hydrogen
% Deuterium

FIG. 3

previously in similar work, was adopted to determine the thermal conductivity at 0° C. of hydrogen, deuterium, and of each mixture, and also the accommodation coefficients of hydrogen and deuterium relative to a platinum surface at 0° C.

The hydrogen was prepared from distilled water, the deuterium from deuterium oxide of 99·95% guaranteed concentration, and the gaseous mixtures from mixtures of the water and deuterium oxide, by passing the vapour over magnesium heated to about 500° C. in an electric furnace. Deuterium was also prepared from the oxide by reaction with metallic sodium in vacuo. The percentage concentrations of the gaseous mixtures were obtained from observations of the densities of the liquid mixtures, using a flotation method.

The values obtained of the thermal conductivity at 0° C. of hydrogen and deuterium were $0{\cdot}000418_2$ and $0{\cdot}000308_0$ cal. cm.$^{-1}$ sec.$^{-1}$ deg.$^{-1}$ C., and of the accommodation coefficient $0{\cdot}29_8$ and $0{\cdot}37_6$ respectively. The values of the thermal conductivities of all the gases together with the percentage concentrations are shown in a table, and also represented by means of a graph.

REFERENCES

Archer 1935 *Phil. Mag.* **7**, **19**, 901.
— 1936 *Nature, Lond.*, **138**, 286
Dickins 1934 *Proc Roy. Soc.* A, **143**, 517.
Farkas and Farkas 1934 *Proc. Roy. Soc.* A, **144**, 467
Gregory 1935 *Proc. Roy. Soc* A, **149**, 35.
— 1936 *Phil. Mag.* **7**, **22**, 257.
Gregory and Archer 1926*a* *Proc. Roy Soc* A, **110**, 91.
— — 1926*b* *Phil Mag.* **1**, 593.
— — 1928 *Proc. Roy Soc.* A, **121**, 285.
— — 1933 *Phil. Mag.* **7**, **15**, 301.
Gregory and Marshall 1927 *Proc. Roy. Soc.* A, **114**, 354
— — 1928 *Proc. Roy. Soc.* A, **118**, 594
Kannuluik 1936 *Nature, Lond.*, **137**, 741
Mann and Dickins 1931 *Proc. Roy. Soc.* A, **134**, 77.
Mann and Newell 1937 *Proc. Roy. Soc.* A, **158**, 397.
Nothdurft 1937 *Ann. Phys., Lpz.*, **28**, 2, 137.
Van Cleave and Maass 1935 *Canad. J. Res.* **12**, 372.
Weber 1917 *Ann. Phys., Lpz.*, **54**, 21, 325.

The theory of pressure-ionization and its applications

By D. S. Kothari, Ph.D. (Cambridge)
Physics Department, University of Delhi

(*Communicated by M. N. Saha, F.R.S.—Received* 14 *December* 1937)

It marked a significant advance in astrophysics when M. N. Saha first showed how the degree of ionization in stellar material could be calculated in terms of its temperature and pressure (or density). In the usual theory of thermal ionization the free electrons are treated as a *classical* perfect gas and, therefore, the theory is applicable only so long as the temperature and density of the material are such that the *free* electrons are *non-degenerate* in the sense of the Fermi-Dirac statistics. In the outer atmosphere of a star the condition of non-degeneracy is always satisfied, but in the interior of the white dwarf stars and the planets (and possibly in the interior of other stars as well) the conditions of temperature and density are such that the free electrons form a *degenerate* gas and their behaviour can no longer be described in terms of classical perfect gas.

In the case of *cold*† matter Saha's theory loses its validity and the degree of ionization in degenerate matter must be investigated on other lines. This investigation of the ionization in degenerate matter is of importance in astrophysics, particularly in researches dealing with the internal constitution of the white dwarf stars and also, as has been recently shown (Kothari and Majumdar 1936*a*, 1936*b*; Kothari 1936), in predicting the maximum radius for a *cold* body.

In this paper we shall deal with the theory of ionization in degenerate matter. In Section 1, by an application of the *virial theorem*, we derive a relation which predicts the degree of ionization in degenerate matter in terms of its density. This relation is obtained on certain simplifying assumptions and we cannot regard it as entirely satisfactory.‡ (All the same, we believe it to be better than any given previously.) In section 2 the results of this paper are compared with those previously obtained. Section 3 deals briefly with some astrophysical applications. The important results are summarized at the end of that section.

---

† The word *cold* is used here in a technical sense. Matter will be referred to as *cold* or *degenerate*, when any *free* electrons present constitute a degenerate gas.

‡ One can consider the present theory to be in the same preliminary stage as the theory of metals immediately after its revival by Sommerfeld.

1.[†] The *virial theorem* states that for an assembly of particles interacting according to the inverse square law of force

$$2T + W = 3pV, \qquad (1)‡$$

where T is the total kinetic energy for all the particles of the assembly, W the total potential energy, V the volume and p the external pressure to which the assembly is subject.

Consider material[§] composed of atoms of atomic weight A and atomic number Z compressed to such an extent that some at least of the outer atomic orbits overlap, and therefore the electrons which occupied these levels are rendered (or squeezed) free—in other words we contemplate conditions such that *pressure-ionization* occurs. We can divide the material into similar spherical cells—each cell containing a nucleus and Z electrons. In general, some of these electrons will be *bound* and some will be *free*. If a denotes the radius of the cell, then a will be connected with the density by the relation

$$\frac{a^3}{\gamma_1}\frac{\rho}{Am_{\mathrm{H}}} = 1, \qquad (2)$$

where m_{H} is the mass of the hydrogen atom and γ_1 is a factor of the order unity. The exact value of γ_1 depends on several factors—particularly the lattice arrangement of the atoms.

In estimating the kinetic energy T we include all electrons, bound as well as free. As is now well known from the investigations of Fermi, Thomas

---

[†] In a previous paper (Kothari 1936) the *virial theorem* has been applied in the form $2T + W + W_G = 0$ to the stellar configuration as a whole, where T, W and W_G are the total kinetic, electrostatic and gravitational energies respectively; and a relation between the radius R and the mass M of the configuration was directly obtained. The aim of the present paper is a different one. Here we are primarily concerned with an investigation of the theory of pressure ionization; the astrophysical application of the theory will be taken up in the last section—the results obtained therein going beyond those of the previous paper.

It may be mentioned that the calculation for the kinetic and electrostatic energies in the present paper follows the lines of the previous paper with necessary formal changes.

[‡] The general result for force varying as the nth power of the distance is

$$2T-(n+1)\,W = 3pV.$$

The sign of the force (attraction or repulsion) is immaterial and it may even vary from particle to particle.

The zero level or state for reckoning the potential energy W is the state when the particles are all dispersed to infinity.

[§] Throughout this paper we shall consider material composed of atoms of one element only, as it does not seem practicable at present to estimate W for a mixture of elements.

and others, the bound electrons in an atom can be treated as forming a degenerate gas, and as the material is *cold* the free electrons will also be degenerate. It will lead to no serious error in estimating T if we assume the Z electrons as uniformly distributed in the cell. The usual formula for the kinetic energy of degenerate electron gas is

$$E_0 = \frac{3}{10}\frac{h^2}{m}\left(\frac{3n^*}{8\pi}\right)^{\frac{2}{3}} N, \tag{3}$$

where n^* is the electron concentration,† N the total number of electrons, m the electron mass and h Planck's constant. Substituting Z for N, and $Z/\gamma_2 a^3$ for n^*, where γ_2 is a factor of the order unity, we obtain for the kinetic energy per cell

$$T' = \frac{3}{10}\frac{h^2}{m}\left(\frac{3Z}{8\pi\gamma_2 a^3}\right)^{\frac{2}{3}} Z, \tag{4}$$

and, therefore, T the total kinetic energy of the assembly is obtained by multiplying T' by the number of cells into which the assembly is divided, i.e. $T = T'\left(\frac{\rho V}{Am_{\mathrm{H}}}\right)$, where ρ is the density and V the volume of the assembly. Substituting for a^3 from (2), we have

$$T = \frac{3}{10}\frac{h^2}{m}\left(\frac{3}{8\pi}\frac{Z\rho}{\gamma_1\gamma_2 Am_{\mathrm{H}}}\right)^{\frac{2}{3}} Z\frac{\rho V}{Am_{\mathrm{H}}}. \tag{5}$$

We shall now obtain an expression for W. It is no simple matter to calculate W accurately, but an approximate value for it can easily be obtained We shall assume that the total potential energy is obtained by multiplying the electrostatic energy of a single cell with the number of cells. Assuming as before that the Z electrons are uniformly distributed in the cell of radius a, the cell having at its centre a nucleus of charge $+Ze$, the potential $U(x)$ at any point inside the cell and distant x from its centre will be, after the cell has been built up to radius x,

$$U(x) = \left(Ze - \frac{4\pi}{3}x^3 n^* e\right)\Big/ x. \tag{6}$$

Therefore the potential energy W' of the cell will be given by

$$-W' = \int_0^a \frac{\left(Ze - \frac{4\pi}{3}x^3 n^* e\right)}{x} 4\pi x^2 n^* e\,dx = \frac{9}{10}\frac{Z^2e^2}{a}, \tag{7}$$

---

† n^* denotes the *total* electron concentration. The *free* electron concentration will be denoted by n.

and hence, eliminating a with the help of (2), we have

$$-W = -W'\frac{\rho V}{Am_{\mathrm{H}}} = \frac{9}{10}Z^2e^2\left(\frac{\rho}{\gamma_1 Am_{\mathrm{H}}}\right)^{\frac{1}{3}}\frac{\rho V}{Am_{\mathrm{H}}}. \tag{8}$$

In the case of degenerate matter the pressure effectively depends only on the free-electron concentration (the heavy particles being non-degenerate make negligible contribution), and is given by the usual expression

$$p = \frac{8\pi}{15}\frac{h^2}{m}\left(\frac{3n}{8\pi}\right)^{\frac{5}{3}}, \tag{9}$$

where n denotes the number of *free* electrons per unit volume It is not to be confused with n^*, which represents the total concentration of bound and free electrons.

Let us now define μ, the *mean molecular weight per free electron*, by the relation†

$$n = \frac{\rho}{\mu m_{\mathrm{H}}}. \tag{10}$$

If the material is r times ionized, i.e. per nucleus there are r free electrons and $(Z-r)$ bound electrons, then

$$n = r\frac{\rho}{Am_{\mathrm{H}}}, \quad \text{and} \quad \mu = \frac{A}{r}. \tag{11}$$

The value of μ gives a measure of the degree of ionization. In the case of singly ionized material $\mu = A$ and for fully ionized material $\mu = A/Z$. Eliminating n between (9) and (10) we have

$$p = \frac{K\rho^{\frac{5}{3}}}{\mu^{\frac{5}{3}}}, \tag{12}$$

where

$$K = \frac{8\pi}{15}\frac{h^2}{m}\left(\frac{3}{8\pi m_{\mathrm{H}}}\right)^{\frac{5}{3}}. \tag{13}$$

K may be called the "degenerate-gas constant".

Substituting for T, W and p in (1) the expressions given by (5), (8) and (12) respectively, and after a little reduction, we obtain

$$\mu = \frac{(A/Z)(\gamma_1\gamma_2)^{\frac{3}{5}}}{\left[1-\frac{(\gamma_1\gamma_2)^{\frac{1}{3}}}{\gamma_1^{\frac{1}{3}}}\left(\frac{\Delta ZA}{\rho}\right)^{\frac{1}{3}}\right]^{\frac{3}{5}}}, \tag{14}$$

---

† It may be noted that μ in our work denotes the mean molecular weight per free electron and not the weight averaged over all the particles, i.e. free electrons and ions. We shall speak of μ defined by (10) as the electron molecular weight.

where Δ stands for

$$\Delta = \left[2.3^{\frac{1}{3}}.\pi^{\frac{5}{3}}\frac{me^2m_{\mathrm{H}}^{\frac{1}{3}}}{h^2}\right]^3. \tag{15}$$

In the above expression γ_1 and γ_2 are factors of the order unity. Their exact values are uncertain, but as a rough approximation, we shall replace them by unity.† Then we have

$$\mu = \frac{\mu_0}{\left[1-\left(\frac{\Delta ZA}{\rho}\right)^{\frac{1}{3}}\right]^3}, \tag{16}$$

where $\mu_0 = A/Z$. This relation predicts the degree of ionization in *cold* matter in terms of its density. If we eliminate ρ between equations (12) and (16) we shall obtain a relation between μ and the pressure p. Thus we see that the ionization in *cold* matter depends only on the *density* or *pressure*, and for this reason it is called *pressure-ionization*.

Let ρ^* denote the density when the material is singly ionized ($\mu = A$), then (16) gives

$$\rho^* = \frac{\Delta ZA}{\left(1-\frac{1}{Z^{\frac{1}{3}}}\right)^3}, \tag{17}$$

and we can write (16) itself in the form

$$\mu = \frac{\mu_0}{\left[1-\left(\frac{\rho^*}{\rho}\right)^{\frac{1}{3}}\left(1-\frac{1}{Z^{\frac{1}{3}}}\right)\right]^3}. \tag{18}$$

In the application of the theory which we shall take up in the last section, equation (18) in certain cases has some advantage over equation (16). The assumptions that we have made to enable us to estimate T and W in an elementary way, though they do not interfere with the underlying principles of the theory, yet render equations (16) and (18) valid only as rough approximations. In the case, therefore, where ρ^* can be estimated from other considerations (for example in the case of the alkali metals we might take it to be the density of the normal metal), we can substitute this value of ρ^* in (18); then, presumably, this equation will give better results than a direct application of (16). This point is also brought out in the next section, where we compare the results obtained in this paper with those worked out previously from a somewhat different point of view.

---

† If we ignore the variation of γ_1, γ_2 with density, then as μ must equal A/Z for $\rho \to \infty$, it follows from (14) that $(\gamma_1\gamma_2) = 1$.

2. In previous papers (Kothari and Majumdar 1931, 1936*b*) we have discussed the theory of pressure-ionization on the crude picture that for material which is on an average r times ionized, the (average) volume available per atomic nucleus is less than the volume of r times ionized atom and more than the volume of $(r+1)$ times ionized atom, that is

$$\begin{bmatrix}\text{Number of}\\ \text{nuclei per}\\ \text{unit volume}\end{bmatrix} \times \begin{bmatrix}\text{Volume of}\\ (r-1)\text{ times}\\ \text{ionized atom}\end{bmatrix} > 1 > \begin{bmatrix}\text{Number of}\\ \text{nuclei per}\\ \text{unit volume}\end{bmatrix} \times \begin{bmatrix}\text{Volume of}\\ r\text{ times}\\ \text{ionized atom}\end{bmatrix}. \tag{19}$$

Under such conditions the material will be on an average r times ionized, for all atomic levels which the outer r electrons can occupy, because of the closeness of packing, have been obliterated. Following Miss Swirles (1933) we can express condition (19) in terms of the successive ionization potentials. Let the outermost electron of the r times ionized atom describe a non-penetrating orbit under an effective nuclear charge Z_{eff}, then the ionization potential which we shall denote by ψ_{r+1} is given by

$$\psi_{r+1} = \frac{1}{2}\frac{e^2 Z_{\text{eff}}^2}{a_{\text{H}}},$$

and the radius of the orbit will be approximately $\dfrac{a_{\text{H}}}{Z_{\text{eff}}} = \left(\dfrac{1}{2}\dfrac{e^2 a_{\text{H}}}{\psi_{r+1}}\right)^{\frac{1}{2}}$, where a_{H} is the radius of the first Bohr orbit for hydrogen $\left(a_{\text{H}} = \dfrac{h^2}{4\pi^2 m e^2}\right)$. The volume of the r times ionized atom can be roughly taken to be that of a sphere of radius $a_{\text{H}}/Z_{\text{eff}}$, and hence, after a little reduction, condition (19) is transformed into the following form·

The material will be on an average r times ionized ($\mu = A/r$), provided the density ρ lies between the limits

$$\rho_1^*\left(\frac{\psi_{r+1}}{\psi_1}\right)^{\frac{3}{2}} > \rho > \rho_1^*\left(\frac{\psi_r}{\psi_1}\right)^{\frac{3}{2}}, \tag{20}$$

where
$$\rho_1^* = \frac{3}{2^{\frac{1}{2}}}\frac{1}{\pi}\frac{\psi_1^{\frac{3}{2}}}{(e^2 a_{\text{H}})^{\frac{3}{2}}} A m_{\text{H}}. \tag{21}$$

ρ_1^* has a simple meaning. It follows from the above relation that for the material to be singly ionized the density should not be less than ρ_1^*. If the density is less than ρ_1^* the material will be on an average less than singly ionized, and if it exceeds $\rho_1^*(\psi_2/\psi_1)^{\frac{3}{2}}$ it will be more than singly ionized.

We shall now proceed to numerical work and consider the following two cases:

(a) The material is assumed to be Iron ($A = 55{\cdot}84$, $Z = 26$).

(b) The material is assumed to be Hydrogen ($A = 1$, $Z = 1$).

We first take the case of iron. The successive ionization potentials have been calculated by Hartree from his method of self-consistent field. Fig. 1 exhibits the results of our calculation. The inequality (20) shows that for a given degree of ionization, that is for a given μ, the density lies between certain (fairly narrow) limits which in the figure are indicated by the curves A and B. The ordinates denote μ and the abscissa $\log_{10}\rho/\rho_1^*$. The portion of curves to the extreme right corresponds to complete ionization of the

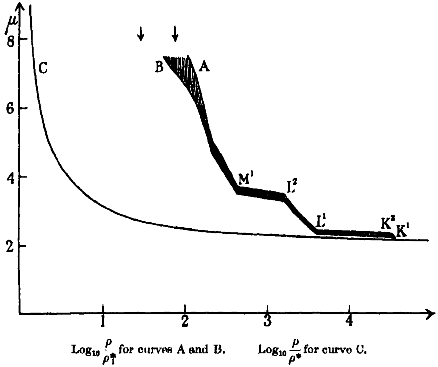


FIG 1 The figure refers to Iron ($A = 55{\cdot}84$; $Z = 26$). The ordinates represent μ, the electron molecular weight, and the abscissa represent $\log\dfrac{\rho}{\rho_1^*}$ for curves A and B, and $\log\dfrac{\rho}{\rho^*}$ for curve C. ρ denotes the density. ρ^* and ρ_1^* denote the densities corresponding to single ionization, as explained in the text.

material to bare nuclei and free electrons. As we move towards the left we meet a bend in the curve A at K^1, which corresponds to the beginning of the formation of the K-shell. At K^2 the K-shell is complete, and at L^1, the formation of the L-shell begins and is completed at L^2. The formation of the M-shell commences at M^1. The further run of the curves is indefinite, as the ionization potentials from ψ_7 to ψ_2 are not known individually, but only their average value (80 V); and for this average value the values of $\log \rho/\rho_1^*$ for the two curves are indicated by two arrows in the figure.

We now turn to the expression for pressure-ionization obtained in Section 1. The curve C in the figure represents μ plotted against $\log_{10}\rho/\rho^*$, as given by equation (18). The discontinuities corresponding to the K, L, M-shells present in the shaded curve AB (the shaded region between the curves A and B will be referred to for brevity as the shaded curve AB) are smoothed out, as is to be expected, in the curve C (which is based on a statistical distribution of the electrons). It will be noticed that leaving aside the region of low ionization the run of the two curves is fairly alike. For low ionization the divergence between them is large; for a given μ the shaded curve AB gives a larger density than curve C.

The density ρ^* corresponding to single ionization in the case of the curve C is found from (17) to be 69·9 g./cm.^{3}, whereas in the case of the shaded curve ρ_1^* the (minimum) density for single ionization is found from (21) to be 63·6 g./cm.^{3} However, for single ionization the density should not be much different from the density of the normal metal (7·86 g./cm.^{3} for iron), and the fact that both ρ^* and ρ_1^* (though agreeing among themselves)† are far removed from this value must serve as a reminder of the rather severe nature of the approximations that occur in the present form of the theory‡. It appears, therefore, that in astrophysical applications of the theory of Section 1 we shall get results in better accord with observation if we take ρ^* in equation (18) to be equal to the density of the normal metal rather than its theoretical value as given by (17). This will be found to be the case in the next section.

---

† Comparing (17) and (21) we find that

$$\frac{\rho^*}{\rho_1^*} = \frac{Z}{\left(1-\frac{1}{Z^{\frac{2}{3}}}\right)^3}\frac{1}{2\pi^3}\left(\frac{\psi^0}{\psi_1}\right)^{\frac{3}{2}},$$

where ψ^0 ($=13·53$ V) is the ionization potential for H atom.

‡ It may be permissible to remark that this comparison between the density of the normal metal and ρ^* (or ρ_1^*) is rather unfair to the theory, as the assumptions made in estimating W, etc. render it particularly inaccurate in the region of low ionization (see also Feinberg 1935).

We now take up the case of pressure-ionization in hydrogen. In this case equation (16) reduces to

$$\mu = \frac{1}{[1-(\Delta/\rho)^{\frac{1}{3}}]^{\frac{3}{5}}}. \tag{22}$$

The following table represents the numerical results obtained from the above equation. The first column denotes the percentage ionization ($100/\mu$), the second the value of μ, and the third and the fourth columns the corresponding values of density and pressure respectively.

TABLE I

Percentage ionization ($100/\mu$)	μ	Density g./cm.^{3}	Pressure dyne/cm.^{2}
0	∞	0·043	0
20	5·00	0·053	$5{\cdot}14 \times 10^9$
40	2·50	0·090	$3{\cdot}89 \times 10^{10}$
60	1·67	0·23	$3{\cdot}63 \times 10^{11}$
80	1·25	1·44	$1{\cdot}26 \times 10^{13}$
100	1·00	∞	∞

The degree of ionization increases with increasing density or pressure.

The relation (20), however, which as already remarked is based on a very crude picture of the phenomena, requires complete ionization for densities above $\rho_1^* = 2{\cdot}68$ g./cm.^{3}, and no ionization for smaller densities. Equation (22) shows that for densities greater than ρ_1^*, the ionization will exceed 84 %, and thus the result following from the crude picture is in reasonably good accord with this equation.

In this connexion it is interesting to note than in a recent paper Wigner and Huntington (1935)—following the lines of the theory of metals given by Wigner and Seitz—have shown the possibility of the existence of metallic hydrogen under pressures of the order of 10^5 atmospheres. (According to Table I, hydrogen under a pressure of 10^5 atmospheres will be about 60 % ionized.) These high pressures are beyond the present laboratory technique, but are of the order that occur in the interior of planets.

In the following section we shall give a brief discussion of the astrophysical applications of the theory.

3. The first astrophysical application of quantum statistics was made by Fowler (1926) in a fundamental paper on the white dwarf stars—stars characterized by a comparatively low luminosity, a high effective temperature and abnormally large mean density of the order of 10^5 g./cm.^{3} This was followed by the work of many investigators, and amongst them Milne's contributions (1932) have been of far-reaching significance and wide appli-

cation. It is now well established that the essential features of the internal constitution of the white dwarf stars can be accounted for by considering their interiors to be composed of *cold* matter.

Let us consider a spherical aggregate of *cold* matter of mass M in equilibrium under its own gravitational forces. Then, its radius R is given by the usual relation (neglecting the effect of relativistic mechanics which is justified so long as the mass M is not larger than that of the sun)†

$$R = \frac{l}{\mu^{\frac{5}{3}}}\left(\frac{\odot}{M}\right)^{\frac{1}{3}}, \tag{23}$$

$$l = \frac{5(\omega_{\frac{3}{2}}^{\circ})^{\frac{1}{3}}K}{2^{\frac{7}{3}}\pi^{\frac{2}{3}}G\odot^{\frac{1}{3}}} = 2{\cdot}79\times 10^{9}\,\text{cm.}, \tag{24}$$

where $\omega_{\frac{3}{2}}^{\circ}$ is a constant (2·1219), characteristic of Emden's solution of Emden's equation of index 3/2, G is the gravitational constant and $\odot$ is the mass of the sun.

The relation (23) involves μ. The question arises: "What value of μ is to be taken in the above formula. Does it depend on M or is it independent of it?" To answer that question we have to bring in the theory of pressure ionization discussed in the preceding sections.

In *cold* matter the degree of ionization is determined essentially by the density or pressure, and if as a first approximation we replace the density ρ in equation (16) by the mean density of the configuration, we have

$$\mu = \frac{\mu_0}{\left[1-\left(\dfrac{\Delta ZA}{3M/4\pi R^3}\right)^{\frac{1}{3}}\right]^{\frac{3}{5}}} \tag{25}$$

Eliminating μ between (23) and (25) we obtain

$$\left.\begin{aligned} R &= \frac{\dfrac{l}{\mu_0^{\frac{5}{3}}}\left(\dfrac{\odot}{M}\right)^{\frac{1}{3}}}{1+\dfrac{l}{\mu_0^{\frac{5}{3}}}\left(\dfrac{4\pi}{3\odot}\right)^{\frac{1}{3}}(\Delta ZA)^{\frac{1}{3}}\left(\dfrac{\odot}{M}\right)^{\frac{2}{3}}} \\ \text{or}\qquad R &= \frac{\dfrac{l}{\mu_0^{\frac{5}{3}}}\left(\dfrac{\odot}{M}\right)^{\frac{1}{3}}}{1+\dfrac{l}{\mu_0^{\frac{5}{3}}}\left(\dfrac{4\pi}{3\odot}\right)^{\frac{1}{3}}\rho^{*\frac{1}{3}}\left(1-\dfrac{1}{Z^{\frac{1}{3}}}\right)\left(\dfrac{\odot}{M}\right)^{\frac{2}{3}}} \end{aligned}\right\}, \tag{26}$$

where for Δ we have substituted from (17) its value in terms of ρ^*.

---

† The formula is quoted from Milne's paper (1932), with the slight change in notation in that his μ is our μm_H and his K is our K/μ.

Equation (26) is fundamental for our purpose, and it leads to several interesting consequences. It shows that as the mass M increases from zero upwards, the radius R at first increases, attains a maximum value, and then decreases approaching zero for $M \to \infty$. If $R_{\max}$ denotes the maximum value for the radius, and M_0 the corresponding mass, then differentiating (26) with respect to M we immediately find

$$R_{\max} = \frac{1}{2}\frac{l^{\frac{1}{2}}}{\mu_0^{\frac{2}{3}}}\left(\frac{3\odot}{4\pi\Delta}\right)^{\frac{1}{6}}\frac{1}{(ZA)^{\frac{1}{6}}} = \frac{1}{2}\frac{l^{\frac{1}{2}}}{\mu_0^{\frac{2}{3}}}\left(\frac{3\odot}{4\pi}\right)^{\frac{1}{6}}\frac{1}{\rho^{*\frac{1}{6}}\left(1-\frac{1}{Z^{\frac{2}{3}}}\right)^{\frac{1}{2}}}, \tag{27}$$

$$\frac{M_0}{\odot} = \frac{l^{\frac{3}{2}}}{\mu_0^{2}}\left(\frac{4\pi\Delta}{3\odot}\right)^{\frac{1}{2}}(ZA)^{\frac{1}{2}} = \frac{l^{\frac{3}{2}}}{\mu_0^{2}}\left(\frac{4\pi}{3\odot}\right)^{\frac{1}{2}}\rho^{*\frac{1}{2}}\left(1-\frac{1}{Z^{\frac{2}{3}}}\right)^{\frac{3}{2}}. \tag{28}$$

We shall now proceed to numerical work. Some assumption has to be made regarding the chemical composition of the material. We shall make here the following alternative assumptions.

(1) The material is assumed to be iron This will be referred to as assumption F.

(2) The material is assumed to be hydrogen. This will be referred to as assumption H.

Under assumption F two further alternatives are possible. As already remarked in the foregoing section we can substitute in (26) for ρ^*, the density corresponding to single ionization, either its theoretical value as given by (17), or identify it with the density of the ordinary (terrestrial) metal (for iron = 7·86 g./cm.^{3}). The first alternative will be referred to as assumption F (*a*), and the second as assumption F (*b*)

Fig. 2 exhibits the results of our calculations. The various curves represent the theoretical mass-radius relation on the different assumptions. The upper curve (HH) corresponds to the case of hydrogen (assumption H). The middle curve $F_B F$ and the lower curve $F_A F$ refer to the case of iron, the former when ρ^* is identified with the density of the metal (assumption F (*b*)) and the latter when ρ^* is given its theoretical value (assumption F (*a*)). For a given value of M, the (M, R) curve fixes R, and using this value of R in equation (25) we obtain the value of μ for the given M. In this way a ladder of μ valves has been put down on each (M, R) curve. An inspection of the figure shows that the ionization increases with increasing mass, and for M comparable to the solar mass the ionization has become (almost) complete.

There is a point with regard to these curves that must be mentioned at the very outset. The equation (23) and so also (26) which is based on it, as already remarked, do not take account of the effect of *relativistic mechanics*.† The effect of relativistic mechanics is negligible for small masses, but becomes appreciable for M comparable to the solar mass, and for larger masses

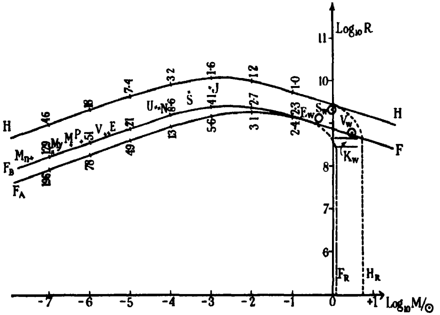


FIG. 2. The curves show the theoretical relation between mass M and radius R (⊙ = mass of the Sun). The curve HH and its modification HH_R due to the effect of relativity are for hydrogen. The curves F_BF and F_AF, and their modified forms F_BF_R and F_AF_R, are for iron—the suffix A stands for the assumption F (*a*) and *B* for the assumption F (*b*).

Mn = Moon.
My = Mercury
V = Venus.
E = Earth.
M = Mars

J = Jupiter.
S = Saturn.
U = Uranus.
N = Neptune.
P = Pluto.

E_W = O_2 Eridani *B*.
S_W = Sirius *B*.
K_W = A.C. 70° 8247.
V_W = Van Maanen's star.

For K_W (the smallest white dwarf observed) the mass is not known. Its radius as estimated by Kuiper (1935) lies between the two horizontal lines indicated in the figure.

the non-relativistic equation (26) breaks down altogether. It has been mentioned above that for $M \gtrless$ ⊙ the theory of pressure ionization predicts the ionization to be complete and hence, as it happens, for a mass for which the relativistic effect has to be taken into account μ equals μ_0. Chandrasekhar (1935) has worked out in detail, taking account of the relativistic effect, the properties (radius, central density, etc.) of the equilibrium configuration of any mass M composed of *cold* matter, and Table III of his paper gives the relation between M and R.

---

† In deriving (23), the "equation of state" used is (12), which holds only so long as the effect of relativistic mechanics is negligible. It may be mentioned, however, that according to the ideas of Eddington (*Relativity theory of protons and electrons*, 1936, § 13·2) this is not so. He holds that (12) is the correct equation in the non-relativistic as well as the relativistic case, and thus according to him (23) and (26) will apply in the relativistic region also.

In the region $M > \odot$ the non-relativistic equation (26) does not apply and we can use Chandrasekhar's table to plot a curve between M and R, μ in his table being put equal to μ_0 in accordance with the theory of pressure-ionization. These curves are shown dotted in the figure. Each dotted curve merges in the corresponding non-relativistic curve in the region of M comparable to and smaller than $\odot$, as the relativistic effect is then negligible. The complete mass-radius relation based on the theory of pressure-ionization and taking into account the effect of relativistic mechanics is shown by the three curves $\mathrm{HH_R}$, $\mathrm{F_BF_R}$ and $\mathrm{F_AF_R}$ on assumption H, assumption F (*b*), and assumption F (*a*) respectively.

We are now in a position to compare the theoretical results with the observed (M, R) values for bodies composed of *cold* matter.

It has been usual to regard the white dwarf stars as composed of cold matter.† We shall make here the assumption that the planets contain a core of cold matter, the size of the core being not much different from the planetary radius. *A priori* there is nothing against our assumption, and at any rate it is worth while to examine its consequences. If, as turns out to be the case, the theoretical results are in reasonable agreement with the observed (M, R) values for planets, then our assumption that the planetary interior is composed of degenerate matter will be placed on almost the same footing as the corresponding assumption for the white dwarf stars.‡

The observed (M, R) values§ for the planets and the white dwarf stars are all shown in fig. 2, and it is indeed interesting to find that the observed values fall in between the extreme theoretical curves $\mathrm{HH_R}$ and $\mathrm{F_AF_R}$ for hydrogen and iron respectively. It will be noticed that, compared to the curves HH_R and F_AF_R, the middle curve corresponding to assumption F (*b*) for iron approximates more closely in the region of terrestrial planets to the run of the curve representing the observed (M, R) values, but the significant fact is that the "observed curve" always lies between the extreme theoretical curves.† The following table gives the values of the maximum radius and the corresponding mass M_0 as estimated from the "observed curve". The theoretical values given by (28) are also given for comparison. The agreement is reasonably satisfactory.

---

† By examining in detail the case of a model white dwarf (Kothari 1933) it has been shown that the thickness of the outer non-degenerate envelope is very small compared to the radius of the degenerate core.

‡ And after all the planets form no uncongenial company with the white dwarfs. In the case of the Earth at any rate the work of Oldham, Guttenberg and others has shown that the interior is a core of molten iron (with possibly a dash of nickel), and therefore is degenerate matter which differs only in the degree of ionization from degenerate matter in the interior of the white dwarfs. The planets, in fine, may be regarded as "black dwarfs"—a term originally due to Fowler for a white dwarf star of zero luminosity.

§ In the case of the Sun and the planets the (M, R) values are taken from H. N. Russell, *The Solar System and its Origin* (1936), p. 10, and for the white dwarfs from Kuiper (1934, 1935).

	R_{max} (cm.)	$M_0/\odot$
From "observed curve"	8×10^9	$1{\cdot}6 \times 10^{-3}$
Theoretical: Assumption H	$12{\cdot}5 \times 10^9$	$1{\cdot}41 \times 10^{-3}$
Assumption F (*a*)	$2{\cdot}77 \times 10^9$	$2{\cdot}79 \times 10^{-3}$
Assumption F (*b*)	$1{\cdot}96 \times 10^9$	$7{\cdot}92 \times 10^{-3}$

It will be observed that the (known) white dwarfs have all masses much larger than M_0 (the mass corresponding to maximum radius), whereas the planets are all smaller than M_0. That the observed white dwarfs are all much more massive than M_0 seems easy to understand, for any white dwarf much smaller than the observed ones would cool too rapidly and would be too faint to be observationally accessible. But there appears to be no immediate explanation of the significant fact *that all (known) planets are smaller than* M_0. In the case of the white dwarf stars, as they happen to be much larger than M_0, the theory of pressure-ionization predicts (see fig. 2) that the material composing them will be (almost) fully ionized.

Another point of interest may be noted.

It will be noticed in the figure that the observed (M, R) values for the two heaviest planets Jupiter and Saturn lie much closer to the Hydrogen curve than to the Iron curves, whereas the terrestrial planets (Mercury, Venus, Earth, Mars) lie nearer to the Iron curves than to the Hydrogen curve. We are thus led to infer that the interior of the outer planets Jupiter and Saturn are in all probability composed of *metallic hydrogen*, whereas the terrestrial planets possess much denser metallic cores.

The main results which have been obtained from the application of the theory of pressure-ionization may be summarized as follows:

(i) The theory predicts that the stellar material in the interior of the white dwarf stars should be fully ionized.

---

† By "observed curve" we mean a smooth curve passing through (or as near as possible to) the observed (M, R) values. This curve is not drawn in the figure, but its general run is evident.

(ii) It predicts the existence of a maximum radius for a *cold* body. The value of this maximum radius is about the same as the radius of the planet Jupiter. *There cannot be a "cold" body (planet or white dwarf) larger in size than Jupiter.*

(iii) The theory shows that the two heaviest planets (Jupiter and Saturn) have cores composed of metallic hydrogen. The terrestrial planets have cores of much heavier metal, possibly iron.

It is a pleasure to thank Professor M. N. Saha for his interest throughout this work. Thanks are also due to Dr R. C. Majumdar and Mr B. N. Singh for friendly discussions.

REFERENCES

Chandrasekhar 1935 *Mon. Not. R. Astr. Soc* **95**, 207-25.
Feinberg 1935 *Phys Z. Sowjet.* **8**, 407–15.
Fowler 1926 *Mon. Not. R Astr Soc.* **87**, 114–12
Kothari 1933 *Mon. Not. R. Astr. Soc.* **93**, 61–90.
— 1936 *Mon. Not. R. Astr Soc.* **96**, 833–43.
Kothari and Majumdar 1931 *Astr. Nachr.* **244**, 65–78.
– – 1936*a* *Nature, Lond.*, **137**, 157–8.
— – 1936*b* *Proc. Nat. Acad. Sci. U.P. India*, **6**, 57-65.
Kuiper 1934 *Pub. Astr. Soc. Pacific*, **46**, 287–90.
-- --- 1935 *Pub. Astr. Soc. Pacific*, **47**, 307–13
Milne 1932 *Mon. Not. R. Astr. Soc.* **92**, 611–43.
Swirles 1933 *Proc. Roy. Soc* A, **141**, 554–66
Wigner and Huntington 1935 *J Chem. Phys.* **3**, 764–70.

The influence of wall oscillations, wall rotation, and entry eddies, on the breakdown of laminar flow in an annular pipe

BY A. FAGE, A.R.C.SC.

(*Communicated by G. I. Taylor, F.R S —Received* 15 *December* 1937)

INTRODUCTION

1. The problem of the breakdown of laminar flow has been much studied, both mathematically and experimentally. The mathematical method of approach is concerned with the prediction of the history of the deviant motion arising from an assumed initial disturbance, superimposed on a selected type of laminar motion and satisfying the equations of viscous fluid flow and the appropriate boundary conditions. A motion is considered to be stable if the deviant motion tends ultimately to vanish, and neutral or unstable if it persists or increases. In general, the disturbance assumed is small and of a form for which the deviant motion can be expressed in terms of linear differential equations, and these restrictions curtail the choice of problem amenable to mathematical treatment.[*]

A review of classical mathematical investigations on the stability of fluid motion has recently been published by Southwell and Chitty (1930). Of these investigations that made by Taylor (1923) for flow in the annular space between two coaxial rotating cylinders, gives an illustration of a successful prediction of instability, experimentally verified. More recent mathematical investigations are those on the stability of flow between parallel planes made by Prandtl (1921, 1931), Tietjens (1925), Tollmien (1929) and Schlichting (1932, 1933*a*, *b*, 1935). These investigations define the mathematical conditions under which the deviant motion grows, and give characteristic numbers, defining the wave-length and frequency detrimental to the maintenance of the laminar state: but these numbers do not necessarily indicate the onset of turbulence. A view commonly held is that breakdown of laminar flow is connected with adverse pressure gradients associated with velocity fluctuations, the magnitude of these gradients depending on the amplitude and frequency of the velocity fluctuations, and increasing with either (Dryden 1931, p. 559; Taylor 1936, p. 308). Further, it has been suggested that it is the amplitude, rather than the frequency, of velocity disturbances which is primarily responsible for the breakdown of laminar flow and the onset of turbulence; and this suggestion has received some support from experiments made by Nikuradse (1933).

---

[*] The problem of large disturbances appears to be, in certain circumstances, within the scope of mathematical treatment (Frazer 1937).

2. The present work deals with experiments* undertaken to obtain information on the breakdown of laminar flow. They deal with the effects of disturbances of known character on a particular type of laminar flow, namely, that for water flowing through a long pipe of annular cross-section. The disturbances considered are those due to axial oscillations of the inner wall of a pipe; those due to oscillations of the inner wall about its axis; and those from discrete eddies introduced into the pipe at its mouth. The first are laminar velocity disturbances unaccompanied by pressure changes; the second are laminar velocity disturbances accompanied by centrifugal pressure gradients; and the third consist of both velocity and pressure disturbances. The work also includes experiments on the breakdown of the axial flow of water in the annular space between two coaxial cylinders, when the outer cylinder is fixed and inner cylinder rotates at a constant angular velocity: and the visual observation of the breakdown of flow near a surface oscillating in a stationary fluid. Finally, theoretical relations for viscous flow at a surface oscillating in a stationary fluid, and for viscous axial flow in an annular pipe, when the inner wall has an axial oscillation and the outer wall is fixed, are derived.

I. WATER SYSTEMS: MEASUREMENT OF PRESSURE DROP

1. The experiments were made for water flowing through vertical pipes of annular cross-section. Each pipe was formed by a smooth brass rod (diameter $2b$) mounted coaxially with a smooth brass tube (diameter $2a$). A faired entry having a length 3 in. and a mouth diameter 4 in. was fitted to the end of the outer tube. Particulars of the pipes used are given in Table I.

2. The Reynolds number of transition from laminar to turbulent flow in a smooth pipe depends, in the absence of imposed disturbances, on the nature of the disturbances in the supply chamber and on those created at the entry. An endeavour was made to obtain steady conditions of flow in

---

* The experiments were made in the Aerodynamics Department of the National Physical Laboratory, and permission to communicate the results was kindly given by the Aeronautical Research Committee.

the entries of the pipes. Steadiness of entry flow cannot be easily and precisely measured; and to make sure that the conclusions drawn were not influenced by the conditions of entry flow, results for different entry systems were obtained.

TABLE I

$m = \frac{1}{2}(a-b)$ = hydraulic mean depth

				Viscous flow	
Pipe	a in.	b in.	m in	$\frac{Q\nu\rho \times 10^3}{a^4 P}$	$\frac{mP}{\frac{1}{2}\rho \bar{u}^2}\left(\frac{\bar{u}m}{\nu}\right)$
I	$0{\cdot}5006_5$	0·4057	0·0475	3·23	5 99
II	$0\ 5006_5$	0·4379	0·0314	0·97	5·99
III	$0{\cdot}5006_5$	0·4095	0·0456	2·88	5 99
IV	$0{\cdot}5006_5$	0·4370	0·0318	1·01	5·99

3. *Entry system A* (fig. 1) This system has a circular inlet tank, divided by three coaxial walls into four compartments—*a*, *c*, *e*, *f*. Water is supplied from a large tank (4 × 2 × 2 ft.), connected to the mains, to compartment *a*, and flows through a ring of holes *b*, into a steadying compartment *c*, a second ring of holes *d*, into the circular compartment *e*, and then through a horizontal wire gauze partition (extending from the circular wall of the circular compartment *e*, to the periphery of the mouth of the inlet) into the pipe inlet. The water-level in the inlet chamber is fixed by the height of the inner wall of the overflow chamber *f*.

Entry system B (fig. 2) Water from the 4 × 2 × 2 ft. tank is supplied to an annular chamber, *a*, fitting on the lip of the faired entry (watertight joint). The water flows continuously through an annular opening in the bottom of this chamber, and then through two wire gauze partitions into the pipe inlet An overflow in the supply tank maintains a constant head in the system.

Entry system C (fig. 3). The pipe inlet is enclosed in a short circular chamber (diameter 5½ in.) leading from an overhead supply tank 2 × 1 × 1 ft. The water in the tank was allowed to stand some time before opening the exit stopcock of the system. The head controlling the flow is maintained constant by a hand device, designed to allow the water-level at the exit to fall with the water-level in the supply tank.

4. The tube forming the outer wall of a pipe was fixed at its exit end to the lid of a circular exhaust tank (fig. 1). The central rod forming the inner wall extended through a watertight gland on the bottom of the tank. The

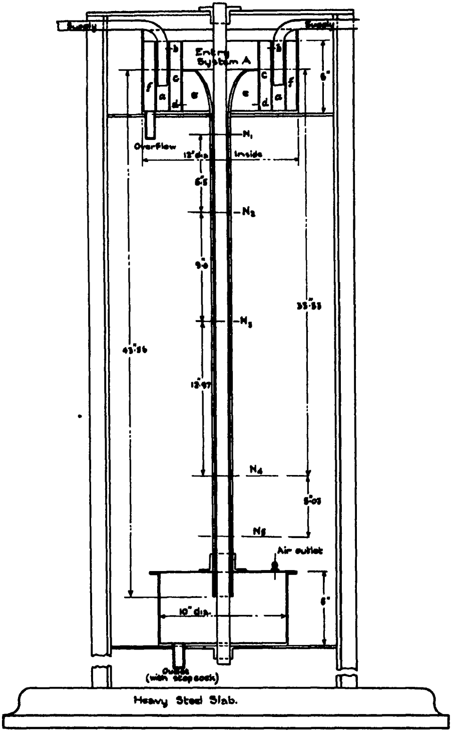


FIG. 1. Entry system A.

flow of water through the system was controlled by a stopcock. The test length of the pipe, N_4N_5 (fig. 1), was 5·03 in., and its upstream cross-section N_4 was 33·53 in. from the mouth of the faired inlet. Four pressure holes, spaced 90° apart, were drilled in each of the sections N_4 and N_5. The measured pressures at the four holes of each section were in very close agreement, and the mean value obtained when they were connected together was taken. The pressure drop $-\partial p/\partial x$, denoted by P, was measured, at low speeds, on Chattock water-carbon tetrachloride gauges (cup centres 13 and 26 in.), and, at high speeds, on a Chattock water-mercury gauge (cup centres 13 in.). The mean velocity of flow $\bar{u}$ was determined from the volume discharge per second, denoted by Q, obtained from the time taken to fill a tank of known volume.

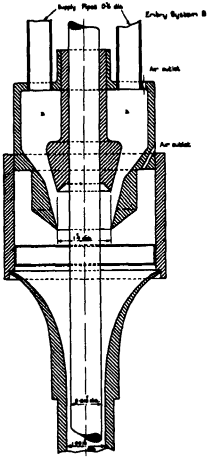


FIG. 2. Entry system B.

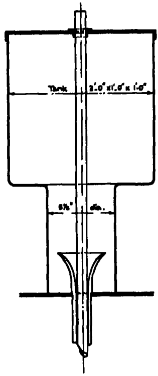


FIG. 3. Entry system C.

5. Values of $mP/\frac{1}{2}\rho\bar{u}^2$ measured at the test section N_4N_5 of pipes III and IV are plotted against $\bar{u}m/\nu$ in fig. 4. The symbols ρ and ν denote the density and kinematic viscosity respectively of the water, and the symbol m the hydraulic mean depth, $\frac{1}{2}(a-b)$. With entry systems A and B the experimental results for both pipes lie closely on the theoretical curve up to a value of $\bar{u}m/\nu = 950$ (approx.): and with entry system C the results fall closely on the theoretical curve up to $\bar{u}m/\nu = 1470$, the value reached with the maximum pressure head available.

6. The distances of the upstream end of the test length N_4N_5 from the entries of the pipes III and IV were 736 and 1055 m. respectively, and the fact that the results for these two pipes, with the same entry conditions, are in agreement, show that statistically uniform conditions of flow are established at the test length. Further, measurements of pressure gradient upstream of the test length were made for pipes I and II over the lengths N_2N_3, N_3N_4, and N_4N_5 (see fig. 1). For each pipe, the values of the non-dimensional coefficient $1000Q\nu\rho/a^4P$ for the three lengths were in close agreement over the entire range of $\bar{u}m/\nu$ up to the critical value. Statistically uniform conditions of flow existed therefore beyond the section N_2 in each of the four pipes.

II. EXPERIMENTS ON THE BREAKDOWN OF LAMINAR FLOW IN AN ANNULAR PIPE WHEN THE INNER WALL OSCILLATES AXIALLY AND THE OUTER WALL IS FIXED

1. To ensure that the disturbances imposed on the water flowing through the pipe arose entirely from the oscillations of the central rod forming the inner wall, and not from extraneous vibrations of the water system, the rod bearings were carried on a stiff framework, erected on a heavy steel slab (fig. 1). The central rod projected at its upper end through the water-level of the upper tank and at its lower end through the bottom of the discharge tank. Two oscillatory systems, *a*, *b*, were used. Each oscillatory system was self-contained and dynamically balanced, the connexion to the central rod being a self-aligning coupling stiff in the axial direction only. Each system gave a true simple harmonic motion to the rod. The amplitudes of oscillation, A, were 0·125, 0·25, 0·50 and 0·64 in.

2. The experiments were made for pipe I with entry system A. The pressure gradient down the test length N_4N_5 was measured at constant values of $\bar{u}m/\nu$, and constant values of the amplitude of the central rod, as

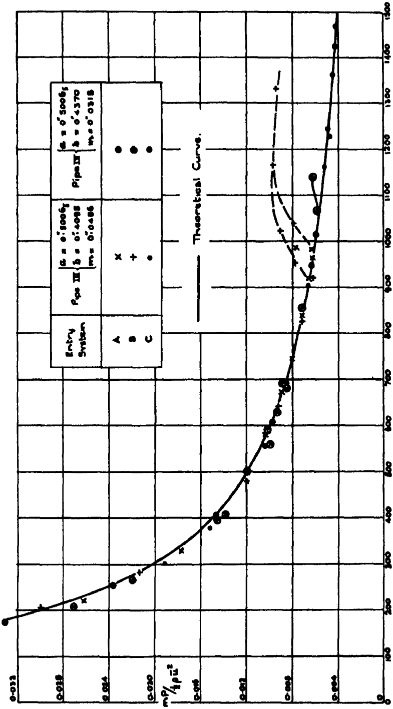

Entry System
A
B
C
Pipe III { a = 0″·5006₅ ; b = 0″·4098 ; m = 0″·0486
Pipe IV { a = 0″·5006₅ ; b = 0″·4370 ; m = 0″·0318
×
+
•
Theoretical Curve.
0·032
0·028
0·024
0·020
0·016
0·012
0·008
0·004
0
mP/½ρū²
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
ūm/ν

FIG. 4

the frequency f was changed over a wide range. The values of $\bar{u}m/\nu$ selected were below the critical value (about 850) with the central rod fixed. Values of ϵ, the ratio of the value of the non-dimensional coefficient $Q\nu\rho/a^4P$ with the rod oscillating to the value (at the same value of $\bar{u}m/\nu$) with the rod fixed were calculated, and a departure of the value of ϵ from unity was taken to be an indication that the flow had departed from the laminar type Some of the experiments were made with the central rod surrounded by a fixed circular tube (internal diameter 1 in.) which cut through the water surface to a depth of about 0·5 in., to eliminate wave formation on the free surface. The results obtained with the guard tube in place were the same as those obtained without the tube.

3. *Results*. The results obtained are given in figs. 5–7, where values of ϵ are plotted against f for constant values of A and $\bar{u}m/\nu$. Each value of ϵ shown is the mean of several repeat values. The departure of ϵ from unity as f increases is very gradual. For each amplitude, the departure or critical value of f, denoted by f_c, is independent, within the accuracy of prediction, of the value of $\bar{u}m/\nu$. The results for $a=0·5$ and 0·64 in. have been plotted together in fig. 7, because the method of prediction is not sufficiently sensitive to allow a distinction to be made between the values of f_c for these two amplitudes.

4. Values of f_c obtained from figs. 5–7 are given in Table II. A Reynolds number of the velocity disturbance due to the oscillation of the inner wall is given by $2\pi fA\lambda/\nu$, where $2\pi fA$ is the velocity amplitude of the wall, and λ is the length $2\pi\sqrt{(\nu/\pi f)}$ (see § VII) Values of λ_c/m, where λ_c is the value of λ when $f=f_c$, and of $2\pi f_c A\lambda_c/\nu$ are included in Table II. It is seen that the value of the critical Reynolds number $2\pi f_c A\lambda_c/\nu$ does not change appreciably over the range of A covered.

TABLE II

Pipe I. $m=0·0475$ in.

(Critical value of $\bar{u}m/\nu$ with central rod fixed = 850.)

A in.	f_c	$\nu \times 10^6$ ft.^{2}/sec.	λ_c/m	$2\pi f_c A\lambda_c/\nu$	Range of $\bar{u}m/\nu$ covered
0·125	19·0	13·5 (April)	0·75	276	234 to 600
0·25	2·7	11·6 (June)	1·85	225	210 to 682
0·57 {0·50	0·9	12·8 (May)	3·37	281	385 to 568
0·64					310 to 550

5. The character of a disturbance from an oscillatory wall of a pipe is illustrated by the velocity curves of fig. 8. These curves, calculated from

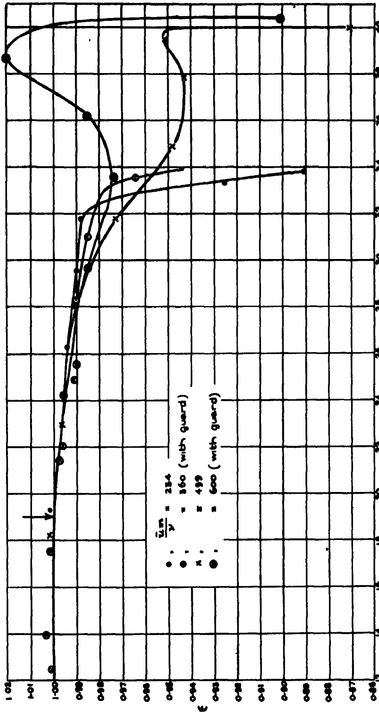


FIG. 5. Pipe I. Oscillatory system (*a*). Amplitude = 0·125 in. $\nu = 13 \cdot 5 \times 10^{-4}$.

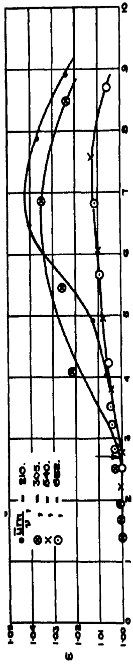


FIG. 6. Pipe I. Oscillatory system (a). Amplitude = 0·25 in. ν = 11·6 × 10⁻⁶.

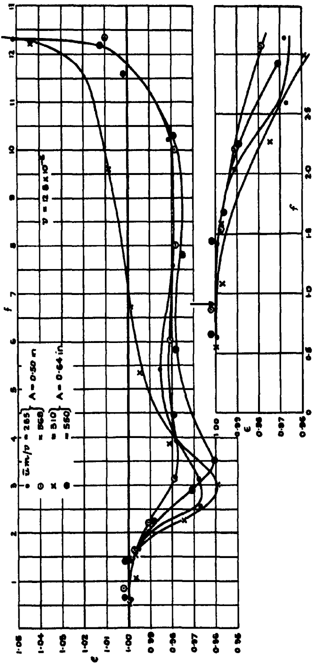


FIG. 7. Pipe I. Oscillatory system (b).

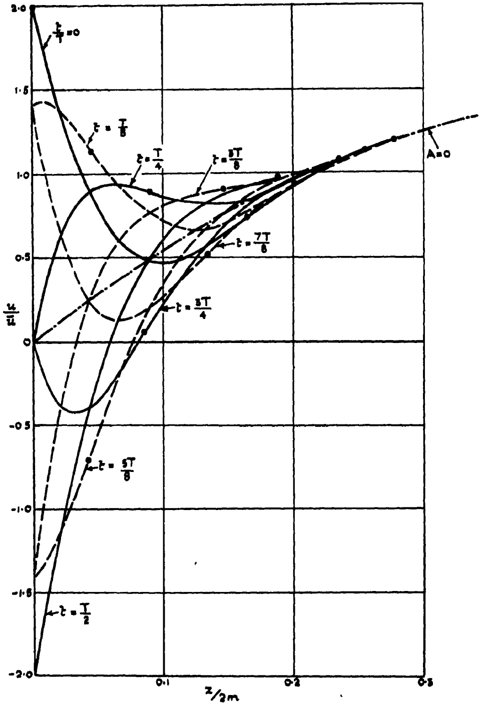

2·0
1·5
1·0
0·5
0
-0·5
-1·0
-1·5
-2·0
$u/\bar{u}$
$t/T = 0$
$t = T/8$
$t = T/4$
$t = 3T/8$
$A = 0$
$t = 7T/8$
$t = 3T/4$
$t = 5T/8$
$t = T/2$
0·1
0·2
0·3
$z/2m$

FIG. 8

the relation 6 of § VII, for the conditions $\bar{u}m/\nu = 200$, $A = 0\cdot12$ in., $f = 19\cdot1$ and $\nu = 11\cdot9 \times 10^{-6}$ ft.^{2}/sec., give the velocity distributions at successive intervals of one-eighth the periodic time, T. The dots on the curves mark the positions of points of inflexion The value of $2\pi fA\lambda/\nu$ for these curves, 282, is very close to the measured critical value. Curves for values of $2\pi fA\lambda/\nu$ below the critical also have points of inflexion, so that the presence of points of inflexion does not constitute a criterion of breakdown of the laminar flow.

III. Experiments on the breakdown of laminar flow in an annular pipe when the inner wall oscillates about its axis and the outer wall is fixed

1. A few experiments were made with pipe I to determine the influence of oscillations of the inner wall about its axis on the breakdown of laminar flow. Two oscillatory systems, I and II, were used. System I oscillated the central rod through a short radial arm coupled to an oscillatory lever. The system was not dynamically balanced, but it ran smoothly, without noticeable vibration of the water system, at frequencies below 25/sec., provided the angular amplitude of the rod, θ, did not exceed 15°. With system II an oscillatory motion was given to the central rod by means of an endless reciprocating wire driving a horizontal pulley fitted to the top

Table III

Pipe I ($m = 0\cdot0475$ in.: $b = 0\cdot4057$ in)

Angular amplitude θ°	$\nu \times 10^6$ ft.^{2}/sec.	f	$\bar{u}m/\nu$	ϵ	$2\pi fA\lambda/\nu$	Remarks
8·9*	15·2	11·4	75	0·998	101	No effect
		9·8	412	0·998	94	
8·9*	15·2	21·5	158	0·997	139	No effect
		21·8	477	0·996	140	
14·7*	15·2	12·2	191	1·002	174	No effect
		12·4	568	1·002	175	
58·0†	13·2	1·25	240	0·999	276	No effect
		2·05	240	1·001	353	
		3·00	240	1·000	427	
68·0†	13·2	1·45	715	1·000	296	No effect
		2·60	715	1·000	396	
149·0†	13·2	1·05	240	0·999	555	
		1·75	240	1·000	716	
		2·20	240	1·012	804	Instability?

* System I. † System II.

of the rod. The system ran smoothly at amplitudes below 150°, and at frequencies below 3/sec. Neither system gave a true simple harmonic motion.

The experimental values of θ, f, and ϵ, and the calculated values of $2\pi f A\lambda/\nu$, where $A = \pi b\theta^\circ/180$ and $\lambda = 2\pi(\nu/\pi f)^{\frac{1}{2}}$, are given in Table III.

2. The value of ξ is unity over the ranges of θ and f covered. These ranges were too limited to allow critical values of $2\pi f A\lambda/\nu$ to be reached, but the results suffice to indicate that the flow remains laminar at values of $2\pi f A\lambda/\nu$ above the critical values measured with the rod oscillating axially. It is likely that the criterion for breakdown depends on the value of $\bar{u}m/\nu$ (see § IV).

IV. THE BREAKDOWN OF LAMINAR FLOW IN AN ANNULAR PIPE WHEN THE INNER WALL ROTATES UNIFORMLY, AND THE OUTER WALL IS FIXED

1. The problem of the stability of the flow of a viscous fluid in the annular space between two infinitely long coaxial cylinders rotating uniformly, with no axial flow, has been solved mathematically, and experimentally verified, by Taylor (1923). When the outer cylinder is fixed and $(a-b)/b$ is small, the relation for the angular velocity of the inner cylinder ω_c at which a critical disturbance, which neither increases nor decreases with time, occurs is

$$\frac{\pi^4\nu^2(a+b)}{2\omega_c^2(a-b)^3b^2} = \beta,$$

where $$\beta = 0{\cdot}0571\left[1-0{\cdot}652\left(\frac{a-b}{b}\right)\right]+0{\cdot}00056\left[1-0{\cdot}652\left(\frac{a-b}{b}\right)\right]^{-1}.$$

This relation* can be recast into the form

$$\frac{2\omega_c bm}{\nu}\left(\frac{a+b}{m}\right)^{-\frac{1}{2}} = \frac{\pi^2}{2\sqrt{\beta}}.$$

2. The present experiments were made to determine the breakdown criterion for water flowing through pipe I (entry system A) under the influence of a pressure gradient parallel to the axis, when the outer wall is fixed, and the inner wall rotates with a uniform speed. Values of ϵ, the ratio of the value of $Q\nu\rho/a^4P$ measured at the test length N_4N_5 when the rod rotates at a uniform speed (ω rad./sec.) to the value when the rod is fixed (same value of $\bar{u}m/\nu$) are plotted in fig. 9, against $2\omega bm/\nu$. For each

---

* β is P in Taylor's paper.

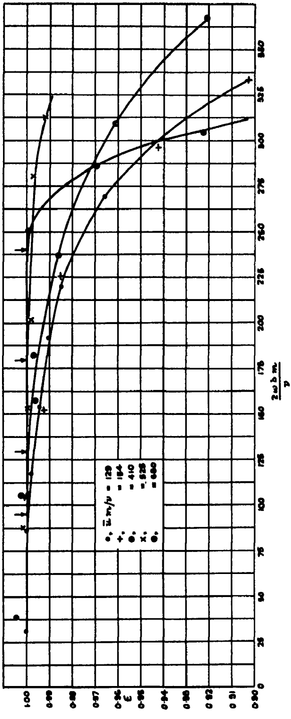

ε
1·00
0·99
0·98
0·97
0·96
0·95
0·94
0·93
0·92
0·91
0·90
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
$\frac{2\omega b m}{\nu}$
•, $\bar{u}\,m/\nu$ = 129
+, = 184
●, = 410
×, = 528
●, = 660

FIG. 9

constant value of $\bar{u}m/\nu$ the value of ϵ (means of several repeat values) lie smoothly on a curve; and the values of $2\omega bm/\nu$ at which the curves depart from the line $\epsilon = 1$ are taken to be the critical values at which departures of flow from the laminar type occur. Values of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/v$, calculated from the critical values of $2\omega bm/\nu$, are plotted against $\bar{u}m/\nu$ in fig. 10. The value of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ increases, at first slowly and then more rapidly, with an increase in $\bar{u}m/\nu$.

For pipe I, β is 0·0491 and Taylor's theoretical value of $\frac{2\omega_c bm}{\nu}\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}$ for $\bar{u}m/\nu = 0$ is 22·3. This value is plotted in fig. 10 and lies on the curve drawn through the experimental values.

3. Recently, Goldstein (1937) has investigated the stability of flow under the influence of a pressure gradient parallel to the axis, with particular reference to the case in which the outer cylinder is stationary and the inner cylinder rotates uniformly.

Goldstein's numerical solutions are for the case when $(a-b)/b$ is small and they cover a low range of um/ν from 0 to 12·92. The values* of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ increase from 20·6 at $\bar{u}m/\nu = 0$ to 22·3 at $\bar{u}m/\nu = 7{\cdot}75$, and then fall, at first slowly, to 21·7 at $\bar{u}m/\nu = 10{\cdot}34$ and then rapidly to 13·8 at $\bar{u}m/\nu = 12{\cdot}92$. The predicted increase of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ with $\bar{u}m/\nu$ is consistent with the experimental indication, but not the rapid fall near $\bar{u}m/\nu = 12{\cdot}92$.

Cornish (1933) has determined, from observations of pressure drop, breakdown criteria for water, when the outer cylinder is fixed and the inner cylinder rotates uniformly. It can be shown that the curve passing through values of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ calculated from Cornish's results for the case $m = 0{\cdot}00918$ in. and $b/m = 429$ lies well above the curve obtained from the present experiments, and on extrapolation gives a value of $2\omega_c bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ for $\bar{u}m/\nu = 0$ which is roughly twice the theoretical value obtained by Taylor. Dr Goldstein has suggested that the phenomenon he considered was different from that observed by Cornish: and the present experiments support this view. Further, the fact that the present

---

* These values have been calculated from values of E_s given in Goldstein's paper.

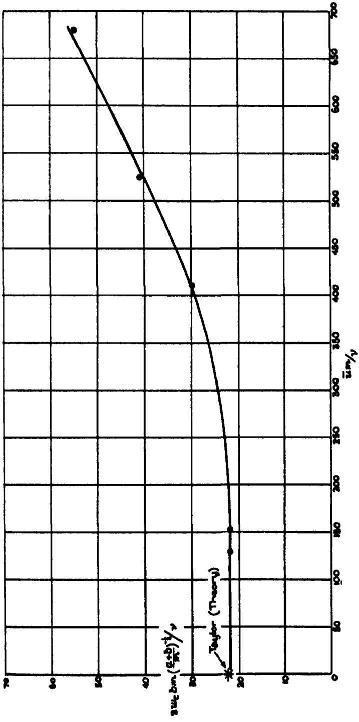

70
60
50
40
30
20
10
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
$\bar{u}m/\nu$
$2\bar{w}_c b_m\left(\frac{a+b}{\pi c}\right)^{\frac{1}{2}}/\nu$
Taylor (Theory)

FIG. 10

experimental values of $2\omega_0 bm\left(\frac{a+b}{m}\right)^{-\frac{1}{2}}\Big/\nu$ and Taylor's theoretical value for $\bar{u}m/\nu=0$ lie on a smooth curve indicates that the present method of observation is reliable.

V. VISUAL EXAMINATION OF THE FLOW NEAR A SURFACE OSCILLATING IN A STATIONARY VISCOUS FLUID

1. The aim of these experiments is to obtain some information on the nature of the breakdown of laminar flow near an oscillating surface. and for simplicity, the case of a surface oscillating in a stationary viscous fluid is taken.

At first, observation was made for a $\frac{1}{8}$ in. flat smooth rectangular brass plate, 16 × 4 in. (chamfered edges) oscillating in the direction of its length on the floor of a water tank, 24 × 5 in., the depth of immersion of the plate below the glass lid being about 1 in. The amplitude of the oscillation was $\frac{1}{4}$ in, and since at the frequencies of observation the disturbance was largely confined to a thin surface layer, it was thought that a clue to the changes of flow which occur as the frequency passes through the critical value might be given by changes of the surface pattern assumed when fine particles, initially in suspension, deposited themselves on the oscillating plate Indeed, fine particles of plaster of Paris arranged themselves in parallel bands normal to the direction of oscillation at frequencies below the critical value predicted from the pipe experiments of § II, and in the neighbourhood of the critical frequency this pattern changed into a rectangular one, with bands in and normal to the direction of oscillation. This significant change was however not convincing for the cause of the parallel-band formation below the critical frequency was not discovered. It was suspected that this formation was a spurious manifestation associated, in some way, with the particles themselves or with an interference flow arising from the free edges of the plate, and this suspicion was proved to be well founded, for heavy particles of silica sand gave the parallel-band formation at frequencies well above the critical frequency

2. Observation of the flows of water and of water-glycerine mixtures in the annular space between two coaxial cylinders when the inner wall had an angular oscillation (outer wall fixed) was then made. The oscillating surface was here continuous in the direction of oscillation, and by the use of minute particles suspended in the fluid to reveal the motion, interference flows of the kind suspected to be present in the flat-plate experiments

were avoided. Fluctuating centrifugal pressures were however introduced by the oscillation.

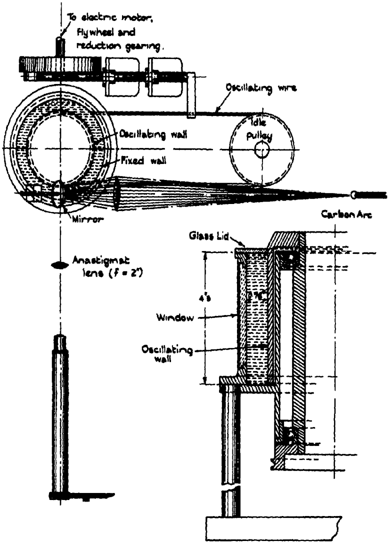


Fig. 11

A general arrangement of the apparatus is shown in fig. 11. Both the inner and outer cylinders ($a = 3{\cdot}218$ in., $b = 2{\cdot}458$ in) were turned from stout gunmetal castings, and the entire apparatus was rigidly constructed to eliminate vibration.

Observation was made normal to the oscillating wall, whose peripheral amplitude was 2 in., and the flow was revealed by minute particles of photographic white ink, illuminated by a bright beam of light. The beam was oblique to minimize convection currents. An image of the field, obtained outside the chamber by an anastigmat lens (focal length 2 in.) was viewed through a microscope, under a magnification of 80 to 1. A fine platinum wire mounted in the focal plane of the eyepiece, and rotatable about the axis of the microscope, was used as a reference line.

3. It was hoped that the advent of the breakdown would be indicated by a marked change in the inclination of the paths traced out by the illuminated particles at the middle of a swing, but observation showed that the inclinations of the paths to the direction of oscillation were, in general, too small for accurate measurement. Changes in flow pattern were however clearly revealed at the ends of a swing, for then the illuminated particles appeared as a cluster of slow-moving bright points of light, which swung upwards, downwards or sideways. Systematic records of these motions were taken over a wide field, for $2\pi fA\lambda/\nu = 234$, 122, 88 and 50. The records for the first two cases are given in Table IV. The motions are specified by the scheme given at the head of the table. Occasionally, a flow changed whilst under observation, and for such flows the letters for the views seen are given

4. The flows at $2\pi fA\lambda/\nu = 234$ and 122 (Table IV) were three-dimensional. Fairly well-defined regions of up flow and of down flow, separated by nodal regions, existed within the fluid, and their existence suggests the presence of ring vortices of the type observed by Taylor (1923) in the breakdown flow between two coaxial rotating cylinders. At $2\pi fA\lambda/\nu = 88$ ($f = 0{\cdot}045$, $\nu = 79 \times 10^{-6}$ ft.^{2}/sec.) the fluid (water-glycerine mixture) appeared to be circulating as a whole, with a down flow at the oscillating wall and an up flow at the fixed wall. At the lowest Reynolds number of observation, $2\pi fA\lambda/\nu = 50$ ($f = 0{\cdot}045$, $\nu = 246 \times 10^{-6}$ ft.^{2}/sec.) the flow approximated very closely to the two-dimensional type, for although some up and down motions were observed (end of swing) they were very weak, and the paths of the particles at the middle of a swing moved in the direction of oscillation.

The breakdown is characterized, then, by a three-dimensional flow, within which nodal regions separating up and down motions exist, and it is not improbable that these motions arise from the presence of ring vortices. The observations are not sufficiently precise to allow a reliable prediction of the value of $2\pi fA\lambda/\nu$ at which the breakdown occurs.

TABLE IV

z denotes the normal distance of the observation point from the oscillating surface.
y denotes the vertical distance from the median horizontal plane (+ve direction upwards).
L denotes left-hand end of swing
R right-hand end.
U up flow: D down flow.
N no deviation from direction of swing (node).
Suffix 1 denotes relatively weak motion.
$\lambda = 2\pi\sqrt{(\nu/\pi f)}$
$2m = 0{\cdot}76$ in. = width of annular gap between the cylinders.

(1) Water $\nu = 11{\cdot}15 \times 10^{-6}$ ft ^{2}/sec.
$f = 0{\cdot}045$, $\lambda = 0{\cdot}67$ in. $= 1{\cdot}76$ m., $2\pi fA\lambda/\nu = 234$

	$z/\lambda = 0{\cdot}075$		$0{\cdot}150$		$0{\cdot}225$		$0{\cdot}300$	
y/λ	L	R	L	R	L	R	L	R
0	U	U	U	U	U	U	U	U
−0·20	N	N	N	N	N	N	D	D
−0·42	D	D	D	D	$U_1 N D_1$	$U_1 N D_1$	U	U
−0·62	D_1	D_1	D	D	D	D	D	D
−0·83	N	N	N	N	$U_1 N D_1$	$U_1 N D_1$	U	U

(2) Water-glycerine mixture $\nu = 41{\cdot}3 \times 10^{-6}$ ft ^{2}/sec.
$f = 0{\cdot}045$; $\lambda = 1{\cdot}28$ in. $= 3{\cdot}3$ m.; $2\pi fA\lambda/\nu = 122$.

	$z/\lambda = 0{\cdot}04$		$0{\cdot}08$		$0{\cdot}12$		$0{\cdot}16$		$0{\cdot}20$		$0{\cdot}24$	
y/λ	L	R	L	R	L	R	L	R	L	R	L	R
0·32	N	N	U_1	U_1	U	U	U	U	U	U	U	U
0·23	D	D	N	N	U	U	U	U	U	U	U	U
0·14	D	D	N	N	U	U	U	U	U	U	U	U
+0·05	D_1	$D_1 N$	U	N	U	U_1	U	U	U	U	U	U
−0·05	D	$D_1 N$	D	$D_1 N$	N	$D_1 N$	U	D	$N U_1$	D	$U_1 N$	D
−0·14	N	N	N	N	N	N	U	N	U	N	U	D
−0·23	D	D	D	D	$D_1 N$	D	N	D	U_1	D_1	U	N
−0·32	D	N	D	D	N	N	N	N	D	N	D	U
−0·42	D	N	D	D_1	D	N	N	N	N	N	U	U
−0·51	D	N	D	N	D_1	N	N	N	U	N	U	U
−0·60	U	N	U	N	N	N	N	N	N	N	U_1	D
−0·69	N	N	N	N	D_1	N	D	D	D	D	D	D
−0·78	N	N	N	D	D	D	D	D	D	D	D	D
−0·87	N	N	N	N	D	D	D	D	D	N	D	U

VI. THE EFFECT OF EDDY DISTURBANCES ON THE BREAKDOWN OF LAMINAR FLOW IN AN ANNULAR PIPE

1. Observations by Schiller (1930, 1934), Naumann (1931, 1935), Kurweg (1933) and Nikuradse (1933) show that the eddies responsible for

the onset of turbulence in pipes and channels arise from the rolling up of surfaces of discontinuity at the entry walls. At a low Reynolds number, the flow at the entry of a pipe is undisturbed and rectilinear. With an increase of Reynolds number undulations of flow appear. At first, these undulations tend to die down within the pipe, but eventually surfaces of discontinuity are formed at the entry walls, followed by a final stage in which these surfaces roll up periodically into discrete eddies, which persist beyond the entry. Below a critical Reynolds number, the energy of the entry eddies is dissipated in the pipe, but above this number, disturbances arising from these eddies persist, and the flow becomes turbulent everywhere

2. The rates at which vorticity were generated at the entries of the pipes and channels used in the above experiments were roughly the same. Systematic variations, over wide ranges, of the rate at which vorticity is generated, and of the frequency, size and strength of the entry eddies, can be obtained by grids of sharp-edged radial plates mounted in the entry normal to the inflowing stream. Further, a grid form of obstruction allows a stream of discrete eddies of known character to be fed into a pipe at Reynolds numbers well below the critical number of the pipe without the grid. The effect of both weak and intense disturbances of this character on the breakdown of laminar flow, well beyond the entry, forms the subject of these experiments.

3 When a grid produces intense disturbances, the influence of a well-designed entry system on these disturbances can be taken to be negligible; but when the grid disturbances are themselves weak, the influence of the entry system cannot be ignored. Anomalies associated with entry flow can be detected by comparison of results obtained for different entry systems, and for this reason two entry systems, B and C (§ 1), were used. Pipe III was used, the critical value of $\bar{u}m/\nu$ with entry system B being about 920, whilst that for system C was above 1470, the highest value of test.

4. Measurements of the pressure drop were made at the test length N_4N_5 of pipe III for ten grids. Each grid was mounted in the junction plane of the faired entry and the pipe, and presented an obstruction in the form of a number of uniformly spaced teeth extending across the width of the pipe. The teeth edges were sharpened by chamfering the undersurface. Details of the grid shapes are given in Table V.

5. The total quantity of vorticity (+ve or −ve) shed in unit time from the sharp edges of the teeth is mnu_s^2, where u_s is the velocity in the vortex

sheet shed from each edge, n is the number of teeth, and $2m$ is the length of a tooth. If the flow in the test length is laminar, the quantity of laminar vorticity (+ve or −ve) passing a cross-section in unit time can be taken as $\pi(a+b)\,(1{\cdot}5u^2)/2$. The ratio of the quantity of vorticity shed at a grid to this quantity of laminar vorticity is $\dfrac{mn}{1{\cdot}125\pi(a+b)}\left(\dfrac{u_s}{\bar{u}}\right)^2$. This ratio becomes $0{\cdot}014n(u_s/\bar{u})^2$ for pipe III. Rough values of $0{\cdot}014n\,(u_s/\bar{u})^2$ for the ten grids, deduced from wind-tunnel experiments on grids, are given in Table V.

TABLE V

Obstruction grids in pipe III.
Length of teeth $= 2m = 0{\cdot}091$ in.

Grid	No. of teeth	Mean width of teeth, w (in.)	Gap/w	$0{\cdot}014n\,(u_s/\bar{u})^2$	Value of $m\bar{u}/\nu$ when $w\bar{u}/\nu = 50$
a^*	20	0·09	0·57	3·8	25
b^*	10	0·18	0·57	1·9	13
c^*	5	0·26	1·23	0·5	9
d^*	4	0·23	2·05	0·25	10
e^*	2	0·20	6·1	0·06	11
$f^\dagger$	2	0·08	17·0	0·06	29
g^*	2	0·08	17·0	0·06	29
$h^\dagger$	2	0·04	35·0	0·06	57
$i^\dagger$	2	0·02	70·0	0·06	114
$j^\dagger$	2	0·01	142·0	0·06	227

* Entry system B. † Entry system C.

The teeth of the grids used in the present experiments can be regarded as small plates normal to the inflowing stream. Experiments on flat plates (Fage and Johansen 1927), circular cylinders (Strouhal 1878; Richardson 1923; Relf and Simmons 1924, Fage 1934), and circular discs (Stanton and Marshall 1930) suggest that eddies will begin to be shed from the teeth at a Reynolds number $\bar{u}w/\nu$ about 50, and at a frequency f, given roughly by the relation $wf/\bar{u} = 0{\cdot}15$. The number, spacing, length, and width of the teeth, influence these numbers, but they should have the right order of magnitude. When $\bar{u}w/\nu = 50$, the pipe Reynolds number $\bar{u}m/\nu$ is $50m/w$. Values of $50m/w$ for the ten grids, given in Table V, show that the grids begin to shed eddies at values of $\bar{u}m/\nu$ ranging from about 9 (grid c) to about 227 (grid j).

6. Fig. 12 shows the values of $mP/\frac{1}{2}\rho\bar{u}^2$ measured at the test length N_4N_5 of the pipe with the four grids (a, c, f, and h) lie close together over the

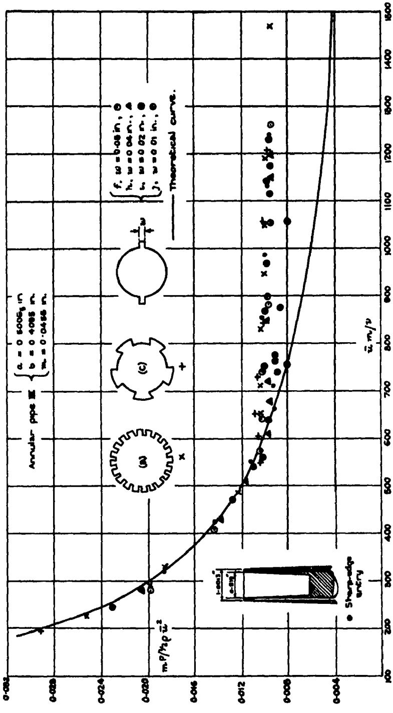

Annular pipe III
a = 0·5005₅ in.
b = 0·4085 in.
m = 0·0456 in.
(a)
(c)
w
f, w = 0·08 in., ○
h, w = 0·04 in., ▲
i, w = 0·02 in., ●
j, w = 0·01 in., ●
Theoretical curve.
1·0003″
0·912″
● Sharp-edge entry
mP/½ρū²
ū m/ν
0·032
0·028
0·024
0·020
0·016
0·012
0·008
0·004
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500

FIG. 12

whole range of $\bar{u}m/\nu$ covered.* The same statistically uniform conditions of flow exist therefore in the test length, in spite of the fact that considerably different amounts of vorticity are shed into the pipe from the entry grids, the vorticity shed from grid *a* being about 60 times greater than that shed from grid *h*. The eddies also differ both in size and frequency; those from grid *a* having diameters about 2·3 times greater and frequencies 2·3 times smaller than those from grid *h*, on the assumption that the eddy diameter is proportional to the tooth width The values of $mP/\frac{1}{2}\rho\bar{u}^2$ depart early, at $\bar{u}m/\nu = 600$, from the theoretical curve for laminar flow. Below this value of $\bar{u}m/\nu$ the pipe walls damp out the entry disturbances; whilst above this value the walls impose the same pattern on the flow at the test length irrespective of the strength of the entry disturbances. The lower critical Reynolds number of the pipe can therefore be taken as 600.

The results for grids *i* and *j* (entry system C) depart later, and less definitely, from the theoretical curve than those for grids *a*–*h* For grid *j*, the departure occurs at about $\bar{u}m/\nu = 750$. The value of $\bar{u}w/\nu$ is then 165, and it is probable that at this small value eddies can just be generated, and that they are shed spasmodically Also, the eddies are small in diameter (about 0·01 in.). It would appear, therefore, that the early breakdown of laminar flow associated with intense eddy disturbances can be caused by very weak entry disturbances, provided they are in the form of discrete eddies.

7. This conclusion was supported by experiments made to measure the effect of obstructions in the form of sharp-edge annular ridges, projecting outwards from the central rod of the pipe. The widths of the ridges were 0·0255, 0·0155 and 0·0055 in. (0·56, 0·34 and 0·12 m.). The first two ridges caused a departure from laminar flow (pipe III, test length N_4N_5, entry systems A and C) at $um/\nu = 600$ (approx.). The values of the Reynolds number uw/ν were then 336 and 204, and it is not unlikely that streams of discrete ring eddies were shed into the pipe. The results for the narrowest ridge (width 0·12 m.) lay on the laminar curve up to $\bar{u}m/\nu = 975$, the highest value of test The value of uw/ν was then 117, and at this low value it is probable that eddies were not shed.

8. Schiller (1932) has suggested that wall roughness has no effect on laminar flow when hu_h/ν, where h is the height of the excrescences and u_h is the velocity in the undisturbed flow at a distance h from the wall, is less than the critical number R_c for the formation of an unsteady wake behind

---

* Grids *b*, *d*, *e* and *g* give same results. Values of $mP/\frac{1}{2}\rho\bar{u}^2$ for a sharp-edged entry to the pipe (no grid) are included in fig. 12.

an obstacle having the shape of the roughness excrescences. The present experiments show that isolated excrescences for which $hu_h/\nu > R_c$ cause, when placed in the mouth of a pipe, a departure from laminar flow at a value of the Reynolds number um/ν lower than that for the departure without the excrescences There is, however, a lower limit of $\bar{u}m/\nu$ below which the flow remains laminar, no matter how large the entry excrescences. Distribution of excrescences of the same kind over the entire surface of the pipe would cause the critical Reynolds number ($\bar{u}m/\nu$) to fall below the lower limit for a smooth pipe, provided, as suggested by Schiller, hu_h/ν were greater than R_c.

VII. THEORETICAL RELATIONS FOR VISCOUS FLUID FLOW IN AN ANNULAR PIPE, AND BETWEEN TWO INFINITE PARALLEL PLANE WALLS, WHEN ONE WALL OSCILLATES AXIALLY AND THE OUTER WALL IS FIXED

1 *Annular pipe*. The equation for viscous fluid flow through a pipe of annular cross-section is

$$\frac{\partial u}{\partial t} - \nu\left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r}\right) = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \tag{1}$$

where u is the velocity in the direction of the axis OX of the pipe, r is the radial distance of a point in the cross-section of the pipe, and $-\partial p/\partial x$ is the pressure gradient down the pipe.

2. For fixed walls the solution (Lamb) of equation (1) is

$$u = \frac{P}{4\mu}\left[(a^2 - r^2) + \frac{(a^2 - b^2)}{\log_e(a/b)}\log_e\frac{r}{a}\right], \tag{2}$$

where P represents $-\partial p/\partial x$

3. *Outer wall, $r = a$, fixed· inner wall, $r = b$, oscillating in the axial direction*. Let the motion of the inner wall be given by $u = A\omega\sin\omega t$. The solution of equation (1) can be written in the form

$$u = u_{A=0} + B\sin\omega t + C\cos\omega t,$$

where $u_{A=0}$ is the solution for $A = 0$ and B and C are functions of r On substitution in equation (1), we have

$$\left[B\omega - \nu\frac{\partial^2 C}{\partial r^2} - \frac{\nu}{r}\frac{\partial C}{\partial r}\right]\cos\omega t - \left[C\omega + \nu\frac{\partial^2 B}{\partial r^2} + \frac{\nu}{r}\frac{\partial B}{\partial r}\right]\sin\omega t$$
$$= \left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right] - \left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right]_{A=0}.$$

This relation is satisfied when

$$\left.\begin{aligned}\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right]C &= \frac{\omega}{\nu}B,\\ \left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right]B &= -\frac{\omega}{\nu}C,\end{aligned}\right\} \qquad (3a)$$

and $$\left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right]=\left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right]_{A=0}. \qquad (3b)$$

A solution of equations (3*a*) in terms of real and imaginary Bessel functions is

$$B = -E_1 \operatorname{ber}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r + E_2 \operatorname{bei}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r - F_1 \operatorname{ker}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r + F_2 \operatorname{kei}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r$$

$$C = -E_1 \operatorname{bei}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r - E_2 \operatorname{ber}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r - F_1 \operatorname{kei}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r - F_2 \operatorname{ker}\left(\frac{\omega}{\nu}\right)^{\frac{1}{2}} r,$$

where the values of E_1, E_2, F_1, F_2 are fixed by the boundary conditions $B=C=0$ when $r=a$, and $B=A\omega$, $C=0$ when $r=b$.

Relation (3*b*) shows that the pressure gradient is not affected by the oscillation of the inner wall.

4. When $\left(\frac{a-b}{a+b}\right)$ is small, as in the experiments of § 2, an approximate relation for the velocity distribution across a pipe, when the inner wall oscillates axially and the outer wall is fixed, is given by the relation for viscous flow between two parallel plane walls, when one wall oscillates, and the other wall is fixed.

5. *Plane walls*. The equation for viscous flow between two parallel infinite plane walls is

$$\frac{\partial u}{\partial t}-\nu\frac{\partial^2 u}{\partial z^2} = -\frac{1}{\rho}\frac{\partial p}{\partial x}, \qquad (4)$$

where the origin is taken in one of the walls, and the axis OZ is perpendicular to them. For fixed walls, the solution is

$$u = \frac{z}{2\mu}[2m-z]P.$$

6. *Wall at $z=0$ fixed: wall at $z=2m$ oscillating in the direction OX.* Let the motion of the wall $z=2m$ be given by $u=A\omega\sin\omega t$. The solution of equation (4) can be written in the form $u=u_{A=0}+B_1\sin\omega t+C_1\cos\omega t$,

where $u_{A=0}$ is the solution when $A=0$, and B_1 and C_1 are functions of z. The substitution of this relation for u in equation (4) gives

$$\left[B_1\omega - \nu\frac{\partial^2 C_1}{\partial z^2}\right]\cos\omega t - \left[C_1\omega + \nu\frac{\partial^2 B_1}{\partial z^2}\right]\sin\omega t = -\frac{1}{\rho}\frac{\partial p}{\partial x} - \left(-\frac{1}{\rho}\frac{\partial p}{\partial x}\right)_{A=0}.$$

This equation is satisfied when

$$\frac{\partial^2 B_1}{\partial z^2} + \frac{\omega}{\nu}C_1 = 0, \tag{5a}$$

$$\frac{\partial^2 C_1}{\partial z^2} - \frac{\omega}{\nu}B_1 = 0, \tag{5b}$$

and

$$\left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right] = \left[-\frac{1}{\rho}\frac{\partial p}{\partial x}\right]_{A=0}. \tag{5c}$$

The boundary conditions are $B_1 = A\omega$ and $C_1 = 0$ at the wall $z = 2m$, and $B_1 = C_1 = 0$ at the wall $z = 0$. The solution of equations (5a), (5b) gives

$$B_1 = R\cos\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z \sinh\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z - S\sin\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z\cosh\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z,$$

$$C_1 = R\sin\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z\cosh\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z + S\cos\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z\sinh\left(\frac{\omega}{2\nu}\right)^{\frac{1}{2}} z,$$

where

$$R = \frac{A\omega\cos m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}\sinh m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}}{\left(\cosh^2 m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}} - \cos^2 m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}\right)}$$

and

$$S = \frac{-A\omega\sin m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}\cosh m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}}{\left(\cosh^2 m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}} - \cos^2 m\left(\frac{2\omega}{\nu}\right)^{\frac{1}{2}}\right)}. \tag{6}$$

7. When $\pi\sqrt{(\nu/\pi f)} < m$, the relation $u = B_1\sin\omega t + C_1\cos\omega t$ for the velocity disturbance can be recast with good accuracy (except near the fixed wall) into the simple form $u = 2\pi f A\exp[-2\pi z/\lambda]\cos\left(\frac{2\pi z}{\lambda} - 2\pi f t\right)$, where λ is the wave length of the disturbance, given by $2\pi\sqrt{(\nu/\pi f)}$, and z is now taken to be the distance from the oscillating wall.

8. In conclusion, the writer wishes to acknowledge his great indebtedness to Mr W. S. Walker, who assisted in the carrying out of the experiments described in the paper.

SUMMARY

1. *Scope of work.* Experiments have been made to determine the effects of disturbances of known character on the laminar flow of water in a long pipe of annular cross-section. The disturbances considered are those due to axial oscillations of the inner wall of the pipe, to oscillations of the inner wall about its axis, and to both weak and intense entry eddies. Experiments on the breakdown due to a uniform rotation of the inner wall (outer wall fixed) have also been made.

The breakdown of flow near a plane surface oscillating in a stationary fluid has been observed.

Theoretical relations for the flow of a viscous fluid through an annular pipe, under the influence of a pressure gradient parallel to the axis, with the inner wall oscillating axially and the outer wall fixed have been obtained.

2. *Conclusions.* The frequency of the axial oscillation of the inner wall when a departure from laminar flow occurs depends on the axial amplitude of the wall and the viscosity of the fluid, and is independent, within the accuracy of measurement, of the velocity of axial flow. The Reynolds number of disturbance, defined as the product of the velocity amplitude at the wall and a length $2\pi\sqrt{(\nu/\pi f)}$ (where f = frequency) divided by the viscosity, at which a departure from laminar flow occurs does not change appreciably over a wide range of amplitude.

The results with the inner wall of the pipe oscillating about its axis suggest that the flow remains laminar up to the critical Reynolds number of disturbance measured with the inner wall oscillating axially.

Visual observation suggests the presence of rotating bands of fluid with their axes parallel to the direction of oscillation at the breakdown.

When the inner wall rotates at a uniform speed (outer wall stationary), the critical rotational speed increases with the axial speed of flow, and the critical number for no axial flow, predicted by extrapolation of the curve drawn through the numbers measured with axial flow, is in close agreement with Taylor's theoretical number

It is shown that the early breakdown of laminar flow associated with intense entry disturbances can be caused by very weak entry disturbances, provided they are in the form of discrete eddies.

REFERENCES

Cornish 1933 *Proc. Roy. Soc.* A, **140**, 227–40.
Dryden 1931 *Rep. Nat. Adv. Comm. Aero., Wash*, No. 392.
Fage 1934 *Proc. Roy. Soc* A, **144**, 381–6
Fage and Johansen 1927 *Proc. Roy. Soc* A, **116**, 170–97
Frazer 1937 *Phil. Mag* **23**, 866–79.
Goldstein 1937 *Proc. Camb. Phil. Soc.* **33**, 41–61
Kurweg 1933 *Ann. Phys., Lpz*, **18**, 193–216.
Lamb "Hydrodynamics", p. 555, 5th ed.
Naumann 1931 *Forsch. Ingwes.* **2**, 85–98.
— 1935 *Forsch. Ingwes.* **6**, 139–45.
Nikuradse 1933 *Z. angew. Math Mech.* **13**, 174–6.
Prandtl 1921 *Z. angew. Math. Mech* **1**, 431–6
— 1931 *Z. angew. Math Mech.* **11**, 407–9.
Rolf and Simmons 1924 *Rep. Memor Aero Res Comm., Lond*, No 917
Richardson 1923 *Proc Phys. Soc.* **36**, 153–65.
Schiller 1930 *Proc 3rd Int Cong Appl. Math.* pp. 226–40.
– 1932 *Handb. exp. Physik*, **4**, 191–2
– 1934 *Z angew. Math. Mech.* **14**, 36–42.
Schlichting 1932 *Nachr Ges. Wiss Göttingen*, 1932 (II), pp 160–98.
– 1933*a* *Nachr Ges. Wiss. Göttingen*, 1933 (II), pp 181–207.
— 1933*b* *Z. angew. Math. Mech* **13**, 171–4.
— 1935 *Nachr Ges Wiss. Göttingen*, **1**, 47–78.
Southwell and Chitty 1930 *Philos Trans* A, **229**, 205–53.
Stanton and Marshall 1930 *Rep. Memor. Aero. Res Comm., Lond*, No 1358.
Strouhal 1878 *Ann. Phys, Lpz.*, **5**, 216.
Taylor 1923 *Philos. Trans* A, **223**, 289–343
– 1936 *Proc. Roy. Soc* A, **156**, 307–17
Tietjens 1925 *Z. angew. Math Mech.* **5**, 200–17
Tollmien 1929 *Nachr. Ges. Wiss Göttingen*, 1929 (I), 21–44

Uranium Z and the problem of nuclear isomerism

By N. Feather, Ph.D., *Fellow of Trinity College,*
and E. Bretscher, *Clerk Maxwell Scholar, Cambridge*

(*Communicated by E. V. Appleton, F.R.S.—Received* 13 *January* 1938)

1. Introduction

It is well known that the idea of the possible isotopic complexity of the elements arose first as the result of investigations into the chemical behaviour of substances produced one from another in the course of successive radioactive transformations. Soon after this idea of isotopy had been generally accepted, Soddy (1917, 1919) suggested a further basis of classification which he believed might also be required in certain cases. According to this suggestion, nuclei which were indistinguishable in respect of charge and mass might still exhibit distinct radioactive properties, or might differ in "any new property concerned with the nucleus of the atom rather than its external shell". He proposed in effect to classify distinct isotopic species as isobaric or heterobaric, depending upon whether the same or different mass numbers had to be assigned to them In particular he suggested that the disintegrations of the branch products thorium C′ and thorium C″, which certainly result in isobaric species of the same atomic number, might in fact give rise to isotopes which were radioactively distinct. This suggestion was tested experimentally by S. Meyer (1918), but was not substantiated. Nevertheless, three years later, uranium Z was discovered by Hahn (1921)—and its chemical and radioactive properties interpreted by him as evidence for nuclear isomerism, in this case with uranium X_2. Subsequent work by Hahn (1923), Guy and Russell (1923) and Walling (1931) confirmed this interpretation. Uranium Z and uranium X_2 were certainly isotopic, and as both were produced, as it seemed directly, from uranium X_1, the most natural assumption to make was that they were also isobaric. No further examples were found, however, and no attempt at a theoretical description of the phenomenon was made for some time. Then, in 1934, Gamow put forward an explanation based upon the idea of an additional fundamental particle, the negative proton, and, from the experimental side, the discovery of artificial radioactivity opened up entirely new possibilities of isotopy. Now, if v. Weizsäcker's (1936) explanation of nuclear isomerism is generally preferred to Gamow's, at least it appears certain that several instances of the phenomenon are open to experimental study (Szilard and Chalmers 1935; Bothe and Gentner 1937; Meitner, Hahn and Strassmann 1937; Snell 1937). For that reason we have thought it important to begin such a study by a more detailed investigation than has hitherto been made of the radiations from uranium Z and the relations between this body and its neighbours in the series. The present paper describes the results of this investigation. So far as they have been taken they are distinctly favourable to v. Weizsäcker's hypothesis; they suggest, therefore, that a precise statement of the hypothesis should be made the basis of our discussion.

2. V WEIZSACKER'S HYPOTHESIS: METASTABLE STATES OF NUCLEI

The hypothesis of v. Weizsäcker is an entirely natural extension of current views regarding the emission of quantum radiations by nuclei, applied specifically to the case of a nucleus in its first excited state. Clearly, from such a state of excitation, the only possible radiative transition is to the ground state of the nucleus—and theory indicates that the probability of this transition depends chiefly upon the amount of energy involved and upon the change in nuclear spin quantum number as between the two states of the system. v. Weizsäcker pointed out that if this energy difference is small, and if the spin change is considerable, very small transition probabilities are to be expected.* He showed that, even when other modes of de-excitation, such as the "mechanical" intervention of an extranuclear electron, were taken into account, the same result held good. Numerical values, given for Z (atomic number) $= 27$ and $Z = 64$, suggest that the effect should not vary much with nuclear charge. For a first excitation level at 5×10^4 e-volts energy, for example, in each case ($Z = 27$ or 64) a spin difference of 5 quantum units† was found necessary to ensure a γ-ray transition probability less than 10^{-5} sec.$^{-1}$ (half-value period greater than 20 hr.). It is interesting to remark that the same suggestion of metastable states of nuclei has been developed, on the basis of the liquid drop model, by Bethe (1937) with numerical conclusions very similar in almost every respect to those quoted above. On this model, however, the dependence of γ-ray lifetime on nuclear charge number is, if anything, even less pronounced than before, and the spin changes necessary to ensure a specified long lifetime are slightly smaller. Bethe also suggests, on a comparison of the theoretical ideas involved in each case, that pairs of nuclear isomers should be about as frequent in nature as pairs of stable isobars having consecutive charge numbers.*

---

* The question of relatively long γ-ray lifetimes was raised in a paper by one of the writers in 1929 (Feather 1929).

† This is an upper limit, only. It was pointed out by Bohr that a high degree of symmetry in the ground state of the nucleus would have the effect of reducing considerably the changes of spin which would be necessary.

The question of the first excited states of the heavy nuclei may be considered from the experimental point of view, also. Here Hulme, Mott, Oppenheimer and Taylor (1936) have already noted certain regularities In the level schemes proposed to account for the γ-rays and the α-particle fine-structure groups from the radioactive elements they have recognized two types, one characterized by a first excited state of quite high energy ($E_1 > 6 \times 10^5$ e-volts), the other by a first excited state not far removed from the ground state of the nucleus. It is possible to extend their list of nuclei having level systems of the latter type (Table I).

TABLE I

Nucleus	RaC	ThC	RaE	RaC″	ThC″
Z	83	83	83	81	81
A	214	212	210	210	208
E_1(e-kV)	52·9	238	47 2	62	40·0

It will be observed that in all the above cases the nuclear charge number (Z) is odd and the nuclear mass number (A) even † The remaining heavy radioactive nuclei of this type are mesothorium 2 and the supposed isomeric pair uranium X_2-uranium Z. Evidently if, following v Weizsäcker, we need to assume that the nucleus of this latter species ($Z = 91$, $A = 234$) has a first excited state very close to the ground state, we shall merely be extending a regularity which is already clearly established. Indeed, there is good reason to suppose that this particular regularity will in fact be further exemplified when the radiations from mesothorium 1 have been successfully studied and information obtained regarding the possible states of excitation of the product nucleus mesothorium 2 ($Z = 89$, $A = 228$). In this case the sequence of disintegrations is as follows:

---

* Seven such pairs of isobars are at present believed to occur amongst stable nuclei (see Bainbridge and Jordan 1936).

† Nuclei for which both Z and A are even provide examples both of low- and high-energy first excited states (thus in the former category occur $_{86}Rn^{222}$, 184e-kV; $_{90}RdTh^{228}$, 58e-kV, $_{88}ThX^{224}$, 86e-kV), but when Z is odd and A even the first excited state of the nucleus appears always to be low-lying. (The statement of Hulme, Mott, Oppenheimer and Taylor that radioactive nuclei of the latter class are those of odd *mass* number is clearly a mistake.) Amongst existing stable nuclei, it will be remembered, this class is remarkable from another point of view—being represented by four species, only (H^2, Li^6, B^{10} and N^{14}).

$$\mathrm{Th} \xrightarrow{\alpha} \mathrm{MsTh_1} \xrightarrow{\beta} \mathrm{MsTh_2} \xrightarrow{\beta} \mathrm{RdTh} \xrightarrow{\alpha} \mathrm{ThX} \rightarrow.$$

Let us consider the changes in nuclear spin quantum number required by existing data regarding these transformations. For the α-disintegrations there is no reason to postulate a change of spin (as between ground states) in either case—in spite of the occurrence of "fine-structure" in the α-radiation from radiothorium.* We assign, therefore, quantum numbers 0 to the ground states both of $\mathrm{MsTh_1}$ and of RdTh. On this assumption the spin of the ground state of $\mathrm{MsTh_2}$ is given from considerations of the β-radiation emitted in the disintegration $\mathrm{MsTh_2} \xrightarrow{\beta} \mathrm{RdTh}$. It will be seen by inspection of fig. 6 (to be discussed in greater detail in § 6) that to explain the partial β-particle spectrum of highest energy in this radiation we must postulate a spin difference of 2 units (or even more) between the ground states of the initial and final nuclei (*vide infra*). It is then necessary to discuss the disintegration $\mathrm{MsTh_1} \xrightarrow{\beta} \mathrm{MsTh_2}$ on the assumption of a spin difference of 2 units between ground states, also. Now up to the present no radiations have been observed in this case, the disintegration constant ($3{\cdot}3 \times 10^{-9}$ sec $^{-1}$) being known only from the rate of growth of radiothorium in initially purified thorium We shall examine, in turn, the assumptions, first that the transformation to which this disintegration constant refers (assuming for simplicity that only one mode of disintegration need be considered) is a "once-forbidden" transition—and then that it is "allowed". On the former assumption, that is assuming transition to the ground state, we conclude (from fig. 6) that the disintegration energy is about $2{\cdot}9 \times 10^5$ e-volts on the latter (transition to some other state) the corresponding energy is less in the ratio 1 : 10. The mere fact that the particle radiation has not been detected (presumably on account of its small energy) is thus strongly suggestive that the latter supposition is correct. But this is seen to imply that the β-particle disintegration of mesothorium 1 results generally in an excited mesothorium 2 nucleus. The additional fact that no strong γ-radiation has been observed is then evidence for the small excitation energy of this state, or for its long lifetime,* or both. Even if it were supposed that transitions to the ground state and to the first excited state of mesothorium 2 were equally probable we should still obtain an energy separation of the two states no greater than $2\cdot3\times10^5$ e-volts on the basis of fig. 6 (partial disintegration energies $2\cdot5\times10^5$ e-volts and $2\cdot0\times10^4$ e-volts, respectively). Empirically, therefore, there appears to be the strongest evidence for low energy excited states in all heavy radioactive nuclei for which Z is odd and A even, and, consequently a good case for the assumption of such a state as a basis of explanation of isomerism in the nucleus uranium X_2-uranium Z. Certain consequences of an explanation on this basis may now be discussed.

---

* Calculation shows that the partial disintegration constants for the low- and high-energy components of this radiation differ by a greater amount than corresponds to the energy difference between them. It is to be supposed, therefore, that the change of spin quantum number occurs between the ground state of radiothorium and the first excited state of the subsequent product, thorium X.

The important consequences of the hypothesis of metastable states were indicated in the original papers of v. Weizsäcker (1936) and Bethe (1937)· they are concerned chiefly with the energy balance in the disintegrations. Here we may consider them, formally, as follows. Let A_0, A_1 be two isomeric nuclei, A_1 representing the higher (metastable) state, of small positive energy ϵ. Let it be supposed that the same type of disintegration occurs with each isomer† and let B be the common product nucleus. Let B_0, B_1, ... denote the ground state and successive excited states of this nucleus. Then, because of the difference in spin between A_0 and A_1, it is unlikely that the transitions $A_0 \rightarrow B_0$, $A_1 \rightarrow B_0$ represent in each case the most probable mode. Rather, it is much more probable that in one case transition to one of the (lower) excited states of B is favoured by the selection rules—or it may be that the radiative transition $A_1 \rightarrow A_0$ is also important. If this is so, then part, at least, of the radiation from A_1 must be of the same quality as that from A_0 (and the particle radiation from a fresh preparation of A_1 is not likely to decay strictly according to the simple exponential law); even if it is not, more than one possibility still remains As most probable disintegration modes we may have

$$(a)\ A_0 \rightarrow B_0,\quad A_1 \rightarrow B_r,$$

or

$$(b)\ A_0 - B_s,\quad A_1 \rightarrow B_0,$$

and in either case if r (or s) $= 1$, and if the first excited state of B is a very low-lying state (excitation energy ϵ'), again differences of disintegration particle energy will not be large ($|\epsilon - \epsilon'|$ or $\epsilon + \epsilon'$). Moreover, if B_1 is very low-lying, it is likely that it will also be metastable: as v. Weizsäcker pointed out we may expect sometimes to find two parallel disintegration series of which corresponding members constitute isomeric pairs. If B_r (or, alternatively, B_s) is not a low-lying state (energy excess given by $\gamma_{r,s}$) then according to either possibility, (*a*) or (*b*), there will be considerable differences between the radiations from A_0 and A_1. If E represents the disintegration energy for the transition $A_0 \rightarrow B_0$, the alternative most probable modes are as follows:

---

* That is, a lifetime not smaller by many orders of magnitude than 6·13 hr., the half-value period of mesothorium 2 for β-transformation—though certain experiments of Widdowson and Russell (1925) make it already unlikely that such a γ-ray lifetime could, in fact, be greater than 3 min.

† Certain experimental evidence (Pool, Cork and Thornton 1937) indicates that this is not always the case -but it does not appear that any fundamental difficulty is introduced if different types of disintegration (e.g. electron and positron emission) have to be taken into account.

(*a*) from A_0, particles of energy E, no γ-rays;
from A_1, particles of energy $E + \epsilon - \gamma_r$, γ-rays of total quantum energy γ_r;
(*b*) from A_0, particles of energy $E - \gamma_s$, γ-rays of total quantum energy γ_s.
from A_1, particles of energy $E + \epsilon$; no γ-rays.

It will be noticed that, when γ-rays are mentioned, the expression "total quantum energy" is employed This is to draw attention to the fact that, since the radiative transition $B_{r,s} \rightarrow B_0$ will generally be "forbidden" on account of the large difference of spin likely to exist between these two states,* the emission of at least two quanta in succession is the most probable de-excitation process.

We may sum up the conclusions of this section, then, by the statement that the radiations from two isomeric radioelements may either be similar or quite dissimilar, according to the energies and angular momenta of the possible excited states of the product nucleus. In this connexion some of the simpler† of the many possibilities have been treated in more detail.

3. EXPERIMENTAL STUDY OF THE RADIATIONS FROM URANIUM Z

Up to the present the only important investigation into the radiations from uranium Z has been that of Walling (1932). Two strong sources (uranium Z corresponding to about 10 kg uranium element) were prepared having very little radioactive contamination (ionization due to the β-particles of uranium X initially only 0·2 and 6·0 % of the total) and the absorption of the β-particles up to 0·65 g./cm.^{2} of aluminium and of the γ-rays up to 12·5 g /cm.^{2} of lead was studied by means of a gold-leaf electroscope. It was concluded, in agreement with the earliest observations of Hahn (1921), that the β-radiation was complex, that is that it could not be described in terms of a single exponential absorption coefficient, and that the γ-radiation could be so described (by means of a mass absorption coefficient of 0·097 cm ^{2}/g. in lead). It was suggested that resolution of the β-radiation into two components for which $(\mu/\rho)_{Al} = 64$ and 19·6 cm.^{2}/g. might be a satisfactory interpretation of the results—if, with an uncovered source, the soft component contributed about four times as much ionization as the harder Beyond 0·33 g./cm.^{2} absorber thickness (of aluminium) no appreciable β-particle effect was recorded No estimate of the intensity of the γ-radiation (in quanta per disintegration) was made.

---

* The initial difference in nuclear spin, as between A_0 and A_1, cannot be expected to be greatly reduced by the two "most probable" (similar) particle disintegrations which are being considered.

† Thus we have supposed, throughout, that transition to the ground state, B_0, was the most probable mode of particle disintegration with one isomer or the other. This need not necessarily be the case.

Our experiments were carried out with the uranium Z obtained from about 2 kg. uranium element, absorption measurements with a tube counter (see Feather 1938) being made with eleven sources prepared by the method of Guy and Russell (1923).* In preparing ten of these sources potassium tantalate equivalent to 20 mg tantalic oxide was used for the precipitation; for the eleventh the equivalent of 200 mg. of oxide was employed. In each case the final precipitate was obtained on a filter paper 1·8 cm. in diameter. Two fusions with potassium hydrogen sulphate usually reduced the uranium X contamination to about 2% initially (i.e. initially about 2% of the β-particles emitted by the source were β-particles from uranium X_2) particularly if a few milligrams of thorium sulphate was added to the melt to assist in this reduction.† After preparation each source was placed in a cavity in a small wooden block of a standard design and covered with a thin foil of mica. Absorption measurements with aluminium were carried out up to a limiting thickness of 1·23 g./cm.^{2} and with lead to a thickness of 9·7 g./cm.^{2}. The absorption of the γ-rays in a block of tungsten of 13·9 g./cm.^{2} was also investigated For the purpose of correction for contamination a complete absorption curve for a source of uranium $(X_1 + X_2)$, with uranium Z in equilibrium, was separately determined‡—and half-value periods of 6·7 hr. (for uranium Z) and 24·5d. (for uranium X_1) were assumed. In the case of each observation corrections were first made for the finite resolving time of the recording system and the natural effect of the counter (see Feather 1938). Afterwards the uranium Z activity through each thickness of absorber was obtained by subtraction and, in the β-particle experiment, was then expressed in terms of the activity through a standard absorber of total thickness* 0·255 g./cm.^{2}. Fig. 1 shows the relative activities obtained in this way for absorber thicknesses greater than 0·16 g./cm.^{2}; the complete absorption curve for all thicknesses being given (logarithmically) by curve A, fig. 2 (β-particle effect only). Fig. 3 gives the logarithmic curve for absorption of the γ-rays in lead. Vertical lines, both in this figure and also in fig. 1, represent limits

---

* Except that, in the original separation of uranium X, iron was employed, instead of thorium, as carrier—and was precipitated as hydroxide.

† The smallest amount of contamination observed after two fusions with $KHSO_4$ was 0·4%, approximately.

‡ This curve (given logarithmically in fig 2, curve *B*) is also required, for purposes of comparison, in the final interpretation of the results (p. 546).

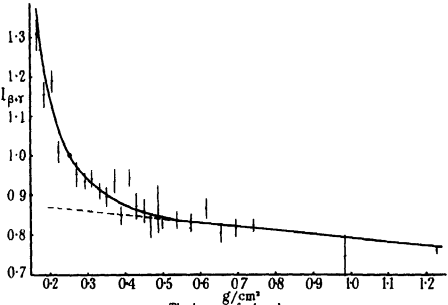


FIG. 1

of probable error based upon the numbers of particles counted For comparison with the results of Walling certain deductions from the curves may be stated at once, leaving until later (§ 6) a more systematic discussion of all the data. Firstly, the β-particle effect is appreciable at least up to 0·5 g./cm.^{2} of aluminium, although it is certainly quite small beyond 0·3 g./cm.^{2}. Secondly, the crude mass absorption coefficient for the γ-rays

---

* Allowance having been made for the mica over the source and closing the counter and for the air included between source and counter. In the γ-ray experiment activities were expressed in terms of the activity through 1·934 g./cm.^{2} of lead, as standard.

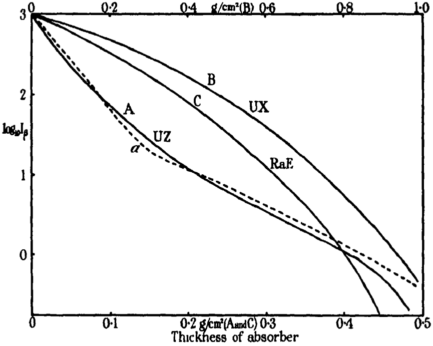

0
0·2
0·4
g/cm²(B)
0·6
0·8
1·0
3
2
1
0
log₁₀I_β
B
C
UX
A
UZ
a
RaE
0
0·1
0·2 g/cm²(A and C)
0·3
0·4
0·5
Thickness of absorber

Fig. 2

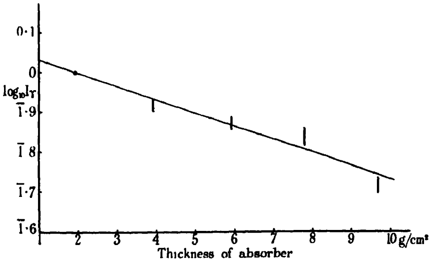

0·1
0
log₁₀Iγ
1̄·9
1̄·8
1̄·7
1̄·6
1
2
3
4
5
6
7
8
9
10 g/cm²
Thickness of absorber

Fig. 3

in lead is $0{\cdot}077 \pm 0{\cdot}006$ cm.^{2}/g., which is considerably less than Walling's value. The crude coefficient for absorption in tungsten, as calculated from the experiments with the single tungsten absorber assuming exponential absorption, is $0{\cdot}073 \pm 0{\cdot}002$ cm.^{2}/g. This is entirely consistent with the other result—and, since it will later be shown that these crude coefficients approximate fairly closely to the true values, it appears that the γ-radiation from uranium Z, as well as the β-radiation, is in fact somewhat more penetrating than was previously supposed

4. THE BRANCHING RATIO

Previous estimates of the branching ratio—that is, the ratio of the probability that an atom of uranium X_1 will disintegrate giving rise to an atom of uranium X_2 to the probability that the disintegration product will instead be an atom of uranium Z—have admittedly been very indirect. Being based upon ionization measurements with relatively weak sources they have been greatly complicated by the considerable difference in penetrating power as between the radiations which were being compared. Direct counting of particles is the only reliable method in such circumstances. For that reason a separate experiment was carried out to determine the branching ratio directly in this way. Three problems were involved: to make sure that the chemical processes employed resulted in the quantitative separation of uranium Z, to correct for the "self-absorption" of the β-particles in the material of the source and, for the comparative measurements, to prepare sources of uranium (X_1+X_2) representing accurately known fractions of the total amount used.

In respect of the first problem, since the precipitation of uranium Z with ten times the usual amount of tantalic acid (cf. p. 536) did not result in the preparation of a source which was any stronger than usual (comparison was made using the γ-radiation, in order to minimize differences due to self-absorption) it was concluded that the standard procedure (using 20 mg. Ta_2O_5) was quantitatively effective.

The standard procedure was, however, modified slightly in order to reduce and make definite the self-absorption of the β-particles of uranium Z in the material of the source. For the main experiments described in the last section the active material had always been prepared as a dried precipitate on a small filter paper, now the final precipitate and filter paper were ignited (with the addition of a few drops of strong nitric acid) and the residue finely powdered and spread uniformly on a circle of smooth white paper inserted in the cavity of a wooden source holder in the usual way. The superficial density of this deposit was very closely 10 mg./cm.^{2}. Fig. 4 gives the results of absorption measurements upon two such sources using thin aluminium foils. With sufficient accuracy a single mass absorption coefficient of 24·7 cm.^{2}/g. may be used in describing the results.* If this coefficient also applies to absorption in the material of the source, calculation shows† that a correction factor of 1·13 is required to take count of self-absorption. Also, inspection of the figure indicates that a factor of 1·38 is necessary to correct for absorption in the standard thickness of 12·8 mg./cm.^{2} of mica and air through which the comparison measurements

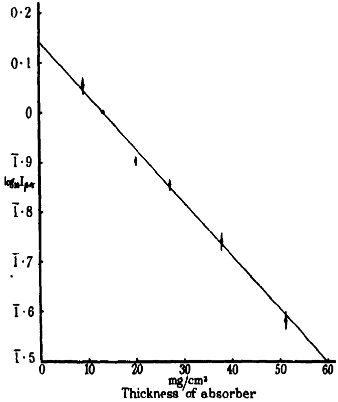


Fig. 4

---

* The initial slope of curve A, fig. 2, corresponds to $(\mu/\rho)_{\text{Al}} = 35{\cdot}0$, the mean slope over the first 0·050 g./cm.^{2} absorption to 30·2 cm.^{2}/g. These values are both greater than that deduced from fig. 4 because the latter includes the γ-ray effect which has been subtracted in fig. 2.

† For a uniform deposit of active material of thickness d the correction factor is $\mu d(1 - e^{-\mu d})^{-1}$.

were made.* Thus a factor of 1·13 × 1·38, or 1·56, was applied to the direct determinations of activity through the standard absorber.

The preparation of sources of uranium ($X_1 + X_2$) for comparison measurements was carried out after division of the original uranium X_1-iron solution into two fractions, *A* and *B*, in the ratio 1 : 19. The larger fraction (*B*) was put aside for subsequent separations of uranium Z, the smaller was made up to 250 c.c. and used 1 c.c. at a time. At any time such a sample (1 c.c) contained, therefore, 1/4750 of the uranium X_1 which remained in the larger fraction. To any such sample about ½ mg. thorium (as nitrate) and 5 mg. iron (as chloride) were added and a precipitate obtained with ammonia. This precipitate was dried and its activity measured in the usual way. The uranium Z was separated from the larger fraction, by the method already described, at a time when it had grown completely to its equilibrium amount. Corrected to this time of separation (and also for absorption, using the data given above) relative activities were found to be as follows: uranium X_2 (from 1 c c. of *A*). uranium Z (from *B*) = 449 : 3200. Thus the branching ratio is 449 × 4750 : 3200, or 666 1. Taking count of statistical errors (not greater than 3 %), of the possibility of chemical losses with such small amounts of material, and of any differential effects due to β-particle reflexion (backwards scattering) in the material of the source support, we may write for this ratio 665 ± 65 · 1, as a final figure An independent estimate based upon a comparison of the γ-ray effects (see § 5) of the uranium Z and $\frac{1}{10}$ of the uranium ($X_1 + X_2$) with which it was in equilibrium gave 580 ± 100 1, in good agreement with the above. When these values are compared with the earlier estimates (350 : 1) based upon ionization measurements, the difference is in the direction to be expected; the softer radiations are given greater weight when ionization methods are employed.

5. THE EFFECTIVE QUANTUM ENERGY AND THE INTENSITY OF THE γ-RAYS

A rough estimate of the intensity of the γ-rays of uranium Z was obtained by comparing the activities of the source, as measured by the counter, first when just sufficient absorbing material was used so that the β-particle effect was completely cut out and, secondly, when no absorber was employed (and correction was made for the self-absorption of the β-particles in the material of the source)—and then determining the ratio of the corresponding activities, $I_{\gamma}/I_{\beta+\gamma}$, for another radioactive substance, concerning which some information regarding γ-ray intensities was available.* Since the success of the comparison depends to a great extent upon the use of a standard substance which emits γ-rays of roughly the same quantum energy as those of the element under investigation, mesothorium 2 was chosen as standard, in spite of the fact that our knowledge of the intensities of the γ-rays from mesothorium 2 is less complete than that concerning radium C or thorium C″.

---

* Through this thickness of material the β-particles of uranium X_1 do not contribute appreciably to the activity of a source of uranium ($X_1 + X_2$); for the β-particles of uranium X_2 a 4 % correction for absorption has to be applied.

First, in order to know more exactly the effective quantum energies of the radiations to be compared, the absorption coefficient for the γ-rays from a preparation of mesothorium 2† was determined under the same conditions as obtained in the main experiment described in § 3. The activity of the source was measured through 0·983 g./cm.^{2} of aluminium, to cut out the β-particle effect, and then after the addition of 1·934 g./cm.^{2} of lead and 13·9 g./cm.^{2} of tungsten, successively Assuming exponential absorption, the following mass absorption coefficients were deduced:

$$(\mu/\rho)_{\text{Pb}} = 0{\cdot}063 \pm 0{\cdot}009 \text{ cm.}^2/\text{g.}; \ (\mu/\rho)_{\text{W}} = 0{\cdot}061 \pm 0{\cdot}001 \text{ cm.}^2/\text{g.}$$

The first of these values is necessarily only rough, however, the more accurate result obtained with the thicker tungsten absorber may be taken to indicate a mass absorption coefficient of 0·065 cm.^{2}/g. in lead. Now the accepted value for this coefficient for the strongly filtered radiation is 0·057 cm.^{2}/g. (Bothe 1924) and the effective value, over the range of absorber thickness here employed, about 0·073 cm.^{2}/g. Thus crude mass absorption coefficients of this order are, with our arrangement, some 0·008 cm.^{2}/g. less‡ than the true coefficients, and we obtain a true coefficient of $0{\cdot}087 \pm 0{\cdot}005$ cm.^{2}/g in lead as best representing all the measurements on the γ-rays of uranium Z described in § 3. Corresponding to this value the effective quantum energy is $0{\cdot}70 \pm 0{\cdot}05 \times 10^6$ e-volts. This must be compared with the known quantum energies 1·65, 1·61, 0·97, 0·915, 0·462 ... $\times 10^6$ e-volts in the γ-ray spectrum of mesothorium 2. Fig. 5 shows roughly how the efficiency of our counter system depends upon the quantum energy of the radiation under investigation.* On the arbitrary scale of this figure the activity to be expected, on the basis of 0·2 quantum per disintegration of mean quantum energy $1{\cdot}63 \times 10^6$ e-volts and 0·3 quantum per disintegration of $0{\cdot}94 \times 10^6$ e-volts energy, is 0·277, per millicurie of mesothorium 2.† Similarly, for 1 quantum per disintegration of uranium Z, of mean quantum energy $0{\cdot}70 \pm 0{\cdot}05 \times 10^6$ e-volts, an

---

* This method has recently been used by Ellis and Henderson (1935) for the annihilation radiation of radiophosphorus (P^{30}) and by Devons and Neary (1937) for the γ-rays of RaC″.

† This source was prepared from the material and by the method described by Feather (1938).

‡ That is, the correction for scattering back into the counter is greater than the correction for obliquity in passing through the absorber, under the relatively poor geometrical conditions of the experiment.

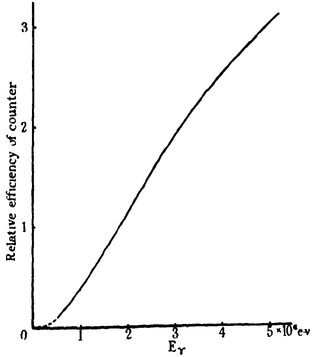


FIG. 5

---

* This curve, which refers only to the electrons liberated in the Compton scattering process, has been obtained by extending the calculations of Richardson and Kurie (1936) for the case of an "infinitely thick" scattering foil. With these authors we have considered just those electrons projected within 10° of the direction of incidence of the scattered quantum—and we have made the reasonable assumption that, with our thin-windowed counter, it is chiefly the electrons so projected from the upper layers of the absorbing foils which are counted when the direct β-particle effect is cut off. That fig. 5 is also fairly well representative of a large number of experiments with thin-windowed counters of different types is suggested by the fact that, whereas Ellis and Henderson (1935) obtained 10 : 1 for the ratio of the γ-ray effects (per millicurie) of thorium C″ and radiophosphorus, our curve gives 9 : 1—on the assumption that only the annihilation radiation is in question with P^{30}.

† This suggestion regarding the spectral distribution of the γ-ray energy of mesothorium 2 is based on the assumption that about 80 % of the total energy given by the heating effect (Rutherford, Chadwick and Ellis 1930, p. 500) is emitted as quanta of the four highest quantum energies given above.

activity of 0·21 ± 0·03 per millicurie is to be expected. The direct experimental result with our counter, by the method described at the beginning of this section, was

$$\text{For mesothorium 2: } I_\gamma/I_{\beta+\gamma} = 1 : 32{\cdot}7 \pm 1{\cdot}8.$$

$$\text{For uranium Z: } I_\gamma/I_{\beta+\gamma} = 1 : 28{\cdot}8 \pm 2{\cdot}0.$$

Thus the number of quanta per disintegration of uranium Z is given by $\frac{31{\cdot}7}{27{\cdot}8} \times \frac{0{\cdot}277}{0{\cdot}21}$, or 1·50 ± 0·25, after estimation of the probable error.* Again we may sum up, by saying that the γ-radiation from uranium Z has, with our arrangement, an effective quantum energy of about $0{\cdot}70 \times 10^6$ e-volts and an intensity quite definitely greater than one quantum per atom disintegrating. This is an important result which will be referred to again in the discussion.

6. DISCUSSION

With the experimental facts established, two main topics for discussion may be distinguished; they are concerned with the mode (or modes) of formation and with the modes of disintegration of the nucleus $_{91}UZ^{234}$. We shall begin, however, with a short statement of the present position regarding the "Sargent diagram", since the discussion of each topic will involve this—and we have already required it (cf p. 533).

As is well known, the empirical relation between disintegration constant and maximum β-particle energy, first pointed out by Sargent (1933), was a relation between total disintegration constant and energy. However, we now know that alternative disintegration modes frequently occur, and partial disintegration constants and the corresponding disintegration energies should therefore be employed (cf. Gamow 1937, p 152). Fig. 6 is an attempt to construct a Sargent diagram using such data in a more systematic manner than has hitherto been done. The partial disintegration constant (in sec.$^{-1}$) and the disintegration energy (in electron volts) corresponding to the most probable mode of disintegration for each† of the β-active bodies of the three series has been used, and an indication has been given, by means of the closed curves‡ surrounding the points on the figure, of the probable degree of uncertainty attaching to these data.* It will be seen that two smooth curves can be drawn reasonably well amongst the points.† These, according to current terminology, are the curves for "allowed" and "once forbidden" transitions—and we shall attribute nuclear spin changes of 0 or 1 and 2 quantum units, respectively, to these transitions (Gamow and Teller 1936). Points corresponding to less probable modes of disintegration are not included in fig. 6 because in general greater uncertainty attaches to them, usually, however, they lie effectively on, or sometimes between, the curves.‡

---

* These limits of error do not take count of inaccuracies in the assumed relative intensities of the γ-rays of mesothorium 2.

‡ Except AcC and $MsTh_1$, for which no data are available.

† The open curves labelled (RaD) and (ThC″) refer to the unobserved transitions to the ground states of the product nuclei in the two cases indicated.

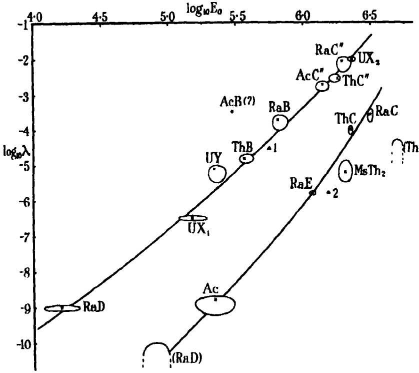


FIG. 6

---

* The uncertainty curves assume different shapes in different cases, and do not lie symmetrically around the "accepted" points, because of the fact that in the absence of further knowledge total disintegration constants have sometimes been used when alternative modes possibly occur, and because all measurements with weak sources—and absorption measurements in particular—are liable to lead to an underestimate of the maximum β-particle energy.

† Except the point for AcB. However, some doubt concerning the β-particle energy still remains (Lecoin 1936)

‡ This—and some of the more considerable of the discrepancies indicated on the figure—does suggest that the assumption of two distinct curves may be an over-simplification of the matter.

Returning then to a consideration of the radiations from uranium Z, we have already noted (§ 3) that at least 0·5 g./cm.^{2} of aluminium is required to absorb the β-particles completely. We conclude (cf. Feather 1930) that the maximum energy of the particles is not less than $1{\cdot}16 \times 10^6$ e-volts. Coupling this value of the energy with the total disintegration constant ($2{\cdot}9 \times 10^{-5}$ sec.$^{-1}$), a point is obtained which on fig. 6 falls almost exactly midway between the curves. The suggestion is that if $1{\cdot}16 \times 10^6$ e-volts is in fact the particle energy corresponding to any mode of disintegration of uranium Z, then certainly it is not the most probable mode. Strong confirmation of this general conclusion is afforded by a further consideration of the logarithmic β-particle absorption curve of fig. 2. It is quite different in form from those referring to the β-particles of uranium X_2 (curve *B*) and radium E (curve *C*), which were also determined. Moreover, since radium E undergoes a "once forbidden" transformation, whereas the disintegration of uranium X_2 is "allowed", the results indicate that these differences cannot be due to differences in the type of disintegration involved. Clearly, the β-radiation from uranium Z must be made up of at least two components with widely different (maximum) energies. An attempt was therefore made to reconstruct curve *A* on the assumption of two components each absorbed according to the simple absorption law represented by curve *C*. The first result to be deduced as the attempt proceeded was that no reconstruction on the basis of any number of components was possible, so long as the most energetic component was assigned the maximum energy of $1{\cdot}16 \times 10^6$ e-volts obtained directly from the absorption curve. The best two-component representation finally devised was that of curve *a*, fig. 2. The components assumed in this case were as follows:

Soft component. absorption limit, 0·20 g./cm.^{2}; intensity, 0·944.

Hard component: absorption limit, 0·70 g./cm.^{2}, intensity, 0·056.

The corresponding maximum β-particle energies are 0·56 and $1{\cdot}55 \times 10^6$ e-volts, respectively. Although the indications are that a better fit still would be obtained on a three-component representation (two soft components and one hard), we shall assume the above analysis as a working hypothesis in the remaining discussions. One result is that the representative points on the Sargent diagram (marked 1 and 2 on fig. 6) now fall much more nearly on the "allowed" and "once forbidden" curves, respectively. We have, therefore, some indication of the spin changes in the assumed alternative modes of disintegration—but we should also notice an apparent difficulty. Whilst the difference between the maximum

β-particle energies ($0{\cdot}99 \times 10^6$ e-volts) is considerably greater than the effective quantum energy of the γ-radiation ($0{\cdot}70 \times 10^6$ e-volts) deduced in § 5, it is not great enough to allow of two successive transitions (and thus nearly two quanta of γ-radiation per disintegration) having this mean energy. However, we shall leave this immediate problem unsolved for the present and turn to a discussion of the modes of formation of uranium Z from uranium X_1. Basing our discussion on v. Weizsäcker's hypothesis of metastable states of low energy, we shall regard that scheme as intrinsically the most likely which, in order to explain the facts, requires us to postulate the metastable state of shortest (γ-ray) life.†

In this connexion there are, broadly speaking, three possibilities: UZ may be the metastable state of the nucleus (UX_2–UZ), in which case it must be formed from UX_1 directly in a rare mode of β-disintegration, or UX_2 may be the metastable state produced in a direct transition from UX_1—and UZ be produced either directly from UX_1, also, or from UX_2 by a (strongly forbidden) γ-transition competing with the much more probable β-disintegration.‡ An asterisk denoting a metastable nucleus (of positive energy ϵ), these three possibilities are represented formally by the schemes (*a*), (*b*) and (*c*), which follow:

(*a*)
$$\begin{tikzcd} & UX_2 \\ UX_1 \arrow[ur, "\beta_1"] \arrow[dr, "\beta_2"'] & \\ & UZ^* \end{tikzcd} \qquad E_{\beta_1} = E_{\beta_2} + \epsilon.$$

(*b*)
$$\begin{tikzcd} & UX_2^* \\ UX_1 \arrow[ur, "\beta_1"] \arrow[dr, "\beta_2"'] & \\ & UZ \end{tikzcd} \qquad E_{\beta_1} = E_{\beta_2} - \epsilon.$$

(*c*)
$$\begin{tikzcd} UX_1 \arrow[r, "\beta"] & UX_2^* \arrow[r, dashed, "\beta"] \arrow[dr, "\gamma"'] & {} \\ & & UZ \end{tikzcd} \qquad E_\gamma = \epsilon.$$

According to (*a*) the lifetime of the metastable state for γ-emission is determined by the amount of β-radiation with the characteristics of the β-radiation from uranium X_2 found in the general particle radiation from uranium Z. If a fraction f of this radiation is observed to be of the same quality as that from uranium X_2, then, approximately, $f\tau_\gamma = \tau_\beta$, where τ_β and τ_γ are the β- and γ-ray lifetimes of the metastable nucleus UZ. Experimentally (from that part of the absorption curve of fig. 1 beyond $0{\cdot}5$ g./cm.^{2} absorber thickness), $f \ngtr \frac{1}{50}$ and $\tau_\gamma \nless 50 \times 6{\cdot}7$ hr. or $\tau_\gamma \nless 335$ hr.

---

† In this way we are choosing the scheme which involves the smallest changes of nuclear spin.

‡ Or both these modes of formation of uranium Z may contribute together.

($\lambda_\gamma \ngtr 5{\cdot}9 \times 10^{-7}$ sec.$^{-1}$). As regards (*b*) and (*c*), clearly (*b*) requires a longer γ-ray lifetime for UX_2^* than does (*c*); according to what has already been said, therefore, we shall discuss only (*c*). Here the γ-ray lifetime is $665 \times \tau$ (UX_2), i.e. $665 \times 1{\cdot}14$ min., or 12·6 hr. ($\lambda_\gamma = 1{\cdot}6 \times 10^{-5}$ sec.$^{-1}$). From these numerical considerations it appears, then, that (*c*) is intrinsically the scheme most likely to be valid in actual fact—and this conclusion is strengthened by evidence from another direction also. For we observe that on the basis of either alternative, (*a*) or (*b*), when the branching ratio is merely the ratio of two β-particle transition probabilities, an "allowed" and a "once forbidden" transition would suffice to explain the branching ratio as determined (665 : 1), whatever reasonable value were assigned to ϵ. What would not be explained, however, would be the very fundamental fact of metastability, since on this basis a difference of spin of 2 units between the two nuclei, UX_2 and UZ, would be the most that could be allowed.

By arguments similar to those already used, we might next treat the problems concerned with the disintegration of uranium Z and uranium X_2, which have been left over, on the basis of v. Weizsäcker's hypothesis But the details are tedious and in some respects the possibilities are many. We shall omit them, therefore, in favour of a final scheme—and we shall content ourselves with pointing out the difficulties which it resolves rather than explaining at length why it appears to us to resolve these difficulties in a more satisfactory manner than is otherwise possible. Our final scheme is as given in fig. 7. Spin quantum numbers are shown on the left and energies (in millions of electron volts, with respect to the ground state of U_{II}) on the right of the energy levels in this diagram, and each transition is also labelled so as to show its type (whether β or γ), its relative probability when it competes with any other transition (upper figure) and its energy (lower figure). Moreover, the vertical energy scale of the diagram is, for sake of convenience, different above and below the horizontal line AA'.

As regards the difficulties resolved, first there is the question of the effective quantum energy and the intensity of the γ-rays of uranium Z. Fig. 7 predicts 1 quantum per disintegration of quantum energy $(0{\cdot}79 - \epsilon) \times 10^6$ e-volts* and 0·944 quantum per disintegration of energy $0{\cdot}99 \times 10^6$ e-volts. This is not, as it stands, completely consistent with the direct results of § 5, but it is evident that to accept a three-component analysis of the β-radiation of uranium Z (p. 546) and a consequent doubling of the level at $(1{\cdot}78 - \epsilon) \times 10^6$ e-volts would improve the agreement. An investigation of the γ-rays of uranium Z by the method of coincidences is being undertaken in the hope of obtaining a critical test of this feature of our scheme. Meanwhile it may be pointed out that if the spin quantum number of the highest excited state of U_{II} is in fact 3 or 4, as shown in fig. 7, then the non-occurrence of the direct radiative transition to the

---

* In fig. 7, and in what follows, the positive energy of the metastable nuclear state is taken as $\epsilon \times 10^6$ e-volts.

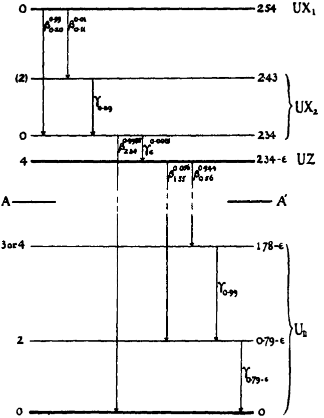


FIG. 7

ground state is just what would be expected. Our second important question concerns the absence of γ-radiation from uranium X_2, in spite of the fact that the disintegration energy ($2 \cdot 34 \times 10^6$ e-volts) is sufficient for excitation of the product nucleus U_{II} in either of the levels shown in fig. 7. This point may be settled at once: the assignment of spin quantum numbers, which has been made in such a way that each of the β-transitions shown is of the type which in fact it is known to be, makes the transitions from UX_2 to the two excited levels of U_{II} once and twice forbidden, respectively. Even the more probable of these transitions, therefore, cannot possess more than one-thousandth of the probability of the transition to the ground state, which is known to occur. Finally, we may mention the difficulty (of long standing) concerning the quantum radiation of $0{\cdot}092 \times 10^6$ e-volts energy, usually assigned to uranium X_1 (Meitner 1923; Hahn and Meitner 1923). We have made a very rough estimate of an upper limit to the intensity of this radiation from various published absorption measurements (and in particular from the data of Richardson (1914)), together with our own value for the total γ-ray activity per millicurie of a source of uranium $(X_1 + X_2)$. Our estimate is 0·01 quantum per disintegration. Now a radiation of this small intensity might quite possibly follow a once forbidden transition of UX_1, and this we have shown tentatively in the figure without attaching great weight to our suggestion. In any case, unless this rare mode of disintegration is in any way involved in the formation of uranium Z (a possibility which we have so far completely disregarded) it is not of any real importance to our main theme.

[*Note added in proof*, 8 *March* 1938. Coincidence experiments on the γ-radiation of uranium Z, carried out in collaboration with Mr J. V. Dunworth, have shown that two γ-ray quanta are emitted "simultaneously" in a large fraction of the disintegrations, as postulated by the level scheme of fig 7. A full account of these experiments will be published shortly.]

7. SUMMARY

The radiations from uranium Z have been examined by the absorption method using a tube counter. The effective quantum energy of the γ-radiation is $0{\cdot}70 \pm 0{\cdot}05 \times 10^6$ e-volts and the intensity $1{\cdot}50 \pm 0{\cdot}25$ quanta per disintegration. An analysis of the β-radiation into continuous spectra with limiting energies 0·56 and $1{\cdot}55 \times 10^6$ e-volts, and intensities in the ratio 17 . 1, is suggested, though it is pointed out that the component of lower energy is probably itself complex. The uranium X_2-uranium Z branching ratio has been determined as $665 \pm 65 : 1$.

On the basis of these results the isomerism of the nuclei UX_2 and UZ is discussed in the light of v. Weizsäcker's hypothesis and a level scheme is put forward which appears to account for all the facts. Reasons are given in support of the conclusion that uranium Z is formed from uranium X_2 in a β-γ branching, rather than from uranium X_1, directly, in a β-β transformation.

REFERENCES

Bainbridge and Jordan 1936 *Phys. Rev.* **50**, 282–96.
Bethe 1937 *Rev. Mod. Phys* **9**, 69–244 (§ 87D).
Bothe 1924 *Z. Phys.* **24**, 10–19.
Bothe and Gentner 1937 *Z. Phys.* **106**, 236–48
Devons and Neary 1937 *Proc. Camb. Phil. Soc.* **33**, 154–63.
Ellis and Henderson 1935 *Proc. Roy. Soc.* A, **152**, 714–23.
Feather 1929 *Phys. Rev.* **34**, 1558–65
— 1930 *Phys. Rev.* **35**, 1559–67.
— 1938 *Proc. Camb. Phil. Soc.* **34**, 115–19.
Gamow 1934 *Phys. Rev.* **45**, 728–9.
— 1937 "Structure of Atomic Nuclei and Nuclear Transformations." Oxford· Clarendon Press.
Gamow and Teller 1936 *Phys. Rev.* **49**, 895–9
Guy and Russell 1923 *J. Chem. Soc* **123**, 2618–31.
Hahn 1921 *Ber. dtsch chem. Ges.* B, **54**, 1131–42
— 1923 *Z. phys. Chem.* **103**, 461–80.
Hahn and Meitner 1923 *Z. Phys.* **17**, 157–67.
Hulme, Mott, Oppenheimer and Taylor 1936 *Proc Roy Soc.* A, **155**, 315–30.
Lecoin 1936 *C R. Acad. Sci, Paris*, **202**, 1057–9.
Meitner 1923 *Z Phys.* **17**, 54–66.
Meitner, Hahn and Strassmann 1937 *Z. Phys.* **106**, 249–70.
Meyer 1918 *S.B. Akad. Wiss. Wien*, IIa, **127**, 1283–96.
Pool, Cork and Thornton 1937 *Phys. Rev* **52**, 380.
Richardson, H. 1914 *Phil Mag.* **27**, 252–6
Richardson, J. R. and Kurie 1936 *Phys Rev* **50**, 999–1006.
Rutherford, Chadwick and Ellis 1930 "Radiations from Radioactive Substances" Camb. Univ. Press.
Sargent 1933 *Proc. Roy. Soc.* A, **139**, 659–73.
Snell 1937 *Phys. Rev.* **52**, 1007–22.
Soddy 1917 *Proc. Roy Inst.* **22**, 117–39.
— 1919 *J. Chem. Soc.* **115**, 1–26.
Szilard and Chalmers 1935 *Nature, Lond.*, **135**, 98.
Walling 1931 *Z. phys. Chem.* B, **14**, 290–6.
— 1932 *Z. Phys* **75**, 425–31
v. Weizsäcker 1936 *Naturwissenschaften*, **24**, 813–14.
Widdowson and Russell 1925 *Phil. Mag.* **49**, 137–40.

The adsorption of vapours at plane surfaces of mica. Part I

By D. H. Bangham and S. Mosallam

(*Communicated by D. L. Chapman, F.R.S.—Received* 20 *January* 1938)

1. Introduction

The measurement of the adsorption of vapours at plane crystalline surfaces is attended by the inherent difficulty of packing into a reasonably small dead-space a sufficient quantity of the solid to give an appreciable effect. The use of powders is open to the objection that the superficial area of the particles cannot be directly measured with accuracy; moreover, their surfaces may contain a network of fine cracks, the effect of which is to multiply the area by an unknown factor (Smekal 1925, Joffé 1928, McBain 1932).

Plates of mica, used first by Langmuir (1918) and later by Bawn (1932), are free from these objections. The peculiar laminar structure of this substance renders it highly unlikely that the sheets of atoms in its exposed cleavage planes are subject to a strong distortional influence such as affects the surface lattices of sodium chloride and similar crystals (Lennard-Jones and Dent 1928). When cleaved, a plate of mica parts along the plane of the potassium atoms; and though the exposed face may comprise a number of "steps", over each of these the surface is "true" to a molecular thickness (W. L. Bragg 1937).

Both Langmuir and Bawn confined their measurements to the region of very low pressures, and in no case was the adsorption found to exceed (or even closely to approach) the value for a unimolecular layer. In this paper an apparatus is described whereby the course of the adsorption could be followed with accuracy from pressures of 0·02 mm. right up to within a few per cent of saturation. The investigation brought to light phenomena of considerable complexity, and showed beyond doubt that multimolecular films are formed at pressures well below saturation.

2. Preliminary experiments

Preliminary experiments with benzene vapour showed the isotherms of this substance to be of the sigmoid form illustrated in fig. 1; they are rather similar to those of water vapour on cellulose and wool, where capillary condensation is usually supposed to play an important part near saturation. In these experiments no attempt had been made to separate the mica strips, as Langmuir had done: tightly wired-up bundles of mica were packed into the adsorption tube so as to economize dead-space. *Prima facie*, therefore, the rapid increase of the adsorption near saturation appeared to be due to capillary condensation between contiguous surfaces of mica. This explanation proved untenable, however, when exactly similar isotherms were found with the mica plates separated from each other by fine wires.

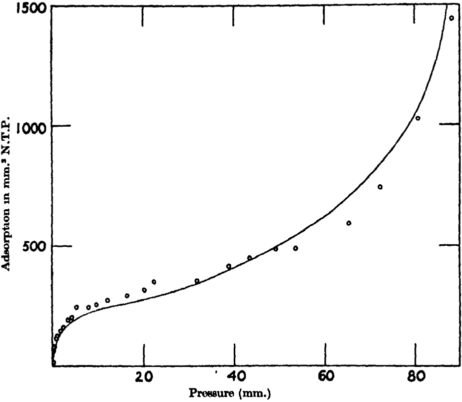


FIG. 1. Isothermals of benzene at 25° C. The continuous curve shows the general course of the isothermal with tightly packed mica strips. The point circles refer to observations where the strips were separated by fine wires.

A comparison of the isotherms obtained with benzene vapour at 25° C. in the two cases will serve to establish this point (fig. 1). The circles refer to measurements with the separated plates, whilst the full curve shows the general trend of the isothermal (ignoring discontinuities) when the plates are packed closely together. Although it is true that the area of the mica was a few per cent greater in the former case, so that the results as they stand are not quite comparable, these measurements show conclusively that the mode of packing the mica is largely immaterial.

Tests for capillary condensation at broken edges. An attempt was next made to discover whether capillary condensation in incipient cleavages at the more or less ragged edges of the plates could be considered the cause of the rapid increase of adsorption near saturation; for although every care had been taken in cutting the strips, and any showing interference colours had been rejected, the edges must undoubtedly have presented a much rougher surface than the cleavage faces.

A mica strip showing well-defined interference patterns due to internal cleavage at an edge was first heated in an evacuated tube, and then exposed to the saturated vapour of benzene, excess of the liquid being present. Even after long standing there was no change in the appearance of the patterns, so that it is certain that capillary condensation in these cleavage spaces did not take place. To speak more generally, we found no evidence that any portion of a clean mica strip is capable of forming a nucleus for the condensation of liquid benzene. The liquid condensed readily on the wire used for suspending the strip, but not at its point of contact with the mica.

To offset against this negative evidence are the facts, brought to light by experiments to be described in a later paper. (1) that after the formation of the first unimolecular layer the heat of adsorption falls abruptly to a value very close to the normal heat of liquefaction; and (2) that towards saturation the quantities of different vapours adsorbed at equal fractions of the saturation pressures represent roughly equal volumes of the normal liquids. Though (as Hückel (1932) has emphasized in the one case, and Polanyi (1933)* in the other) neither of these relations is in actual agreement with the theory of capillary condensation, their approximate validity is so suggestive of the filling up of cavities with liquid phase that the possibility cannot lightly be disregarded. Further experiments were therefore undertaken.

Spreading experiments in the presence of saturated and supersaturated vapours. A plate of mica was supported horizontally in an all-glass system, and, after being given the same vacuum heating as the mica used in the adsorption experiments, was exposed to the saturated vapour of benzene. Drops of liquid benzene were then caused to fall on its surface from the end of a fine capillary. These drops did not spread in the way they should if the surface had been covered with a film of liquid benzene. They flattened, it is true, under the influence of the kinetic energy of their fall, and their contours were often irregular. But it often happened that a second drop, following the first, would roll the latter back from the centre leaving, for the time, a clear space at the point of impact. The behaviour of the drops was rather similar to that of water drops on a slightly greasy surface. The angle of contact could not be estimated on account of the irregular contours of the drop; all that can be stated with certainty is that it was less than 90° and much greater than 0°. Other pure liquids (methyl alcohol, carbon tetrachloride, acetic acid) behaved very similarly to benzene, as also did such liquid mixtures as were tried.

---

* Even if the contact angles were zero for a number of liquids, the relative pressures at which equal volumes of those would be condensed by capillarity are not equal, but would be smaller in the case of liquids of high surface tension and large molecular volume. On account of the broken form of the isotherms obtained with mica, and the rapid increase of adsorption near saturation, a detailed comparison of the data with the requirements of the capillary condensation theory would be difficult, and of doubtful value; it is sufficient to state that the deviations from the "equal volume" rule are often in the sense opposite to that predicted on the basis of this theory, zero contact angles being assumed.

In a further set of experiments, freshly split surfaces of mica were exposed to dry, drop-free air supersaturated with the vapour of benzene and other organic liquids as it issued from a jet into the open. Provided the degree of supersaturation was not too great (it was still great enough to cause immediate bulk condensation on a slightly smoked glass slide), there was no condensation of liquid observed. Instead there appeared on the mica a film rendered visible by colours which we believe to be due to interference On placing small drops of liquid benzene on these coloured films, the drops flattened to an extent depending on the degree of supersaturation of the vapour, but they did not merge their identity with that of the films. We regard these experiments as proving. (1) that the polymolecular films known to exist from the adsorption measurements are situated at the ordinary cleavage surfaces of the plates, and are not concentrated only at the edges; and (2) that these films have properties which differentiate them from the normal bulk liquids.

Obreimoff's experiments. It may be objected, in regard to the experiments last described, that since the mica was for a short time exposed to air after cleavage, the adsorbing surface would probably be not that of mica itself, but a film of adsorbed moisture. Thus Obreimoff (1930), who measured the surface tension of mica by splitting it with a wedge and noting the curvature near the edge of the split, concluded that whilst a freshly cleaved surface in air has a surface tension of 1500 dynes/cm., that of a surface cleaved in a good vacuum is as high as 20,000 dynes/cm.

A rough calculation based on Gibbs's adsorption equation shows that if 10 A^2/molecule be allowed for the water in the monolayer, in order to cause a surface tension lowering of 18,500 dynes/cm., the layer would need to be stable to a 10^{104}-fold reduction of pressure at ordinary temperatures. The film would withstand the most drastic vacuum heating, and it can safely be inferred that if the peculiarities of behaviour towards supersaturated vapours shown by mica freshly split in air are due to a film of adsorbed water. this water film was also present on the degassed mica used for the adsorption measurements with benzene and other vapours. It is of interest to note, in connexion with the possible rôle of a water film, that a jet of air, if even very slightly supersaturated with benzene vapour, gave *immediate* condensation when played on the surface of ordinary water; also that supersaturated water vapour behaved towards freshly split mica in much the same way as has been described for benzene and other organic liquids.

3. Apparatus and materials

The apparatus comprised. (1) a supply bulb containing the experimental liquid, (2) a train for its fractionation under vacuum conditions, (3) the pumping system which included pentoxide tubes and a McLeod gauge reading to 2×10^{-5} mm., (4) the adsorption vessel containing the mica, and (5) a combined burette and pressure gauge. Only the two last require special description.

The measuring system and connecting tubes are sketched in fig. 2. It will be noted that the former has neither taps nor ground-glass junctions, and is designed for complete immersion in a water thermostat. *AAAA* is the thermostat tank, and the frame *BBBB* serves to support the adsorption vessel *C* and the gauge burette *DEFGH*. These two vessels are connected at the inserted join *F*.

The gauge burette comprises: (1) the four bulbs on the limb *DE*; each was of about 15 c.c. capacity, and they were separated by short lengths of tubing provided with etched marks at the positions shown; (2) the bent capillary at *E*, of 1·5 mm. bore; (3) the uniform gauge tube *EFGH* of 8 mm. bore; this had seven finely etched scale marks at positions 0, 1, 2, ..., 6, and led directly to the McLeod gauge and pumping system.

In measuring a quantity of vapour prior to its introduction to the adsorption vessel, the mercury in the gauge tube was first raised above the inserted join at *F*, control being effected through the tap *K*. The mercury was prevented from entering the mica tube by the operation of the float valve *L*. The vapour was then introduced through the cut-out *M*, imprisoned between mercury surfaces set to the etched marks in *DE* and *EF*, and its pressure recorded with the aid of a cathetometer reading to 0·01 mm.

After the distribution of the vapour to the adsorption vessel via the inserted join *F*, its pressure was recorded at intervals until no further change could be observed. The whole procedure was then repeated.

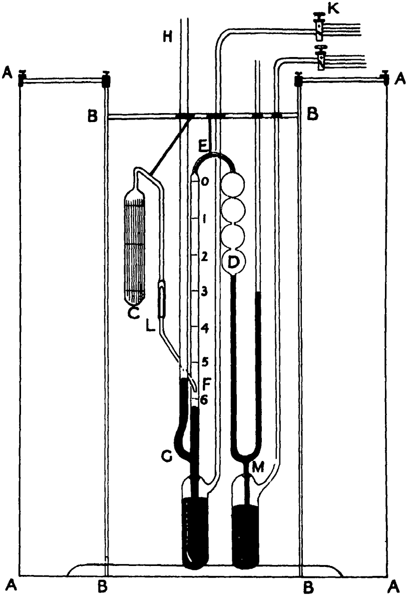


Fig. 2.

The gauge burette is so constructed that the pressure can either be read directly, as was necessarily the case when approaching saturation, or *indirectly*, making use of the principle of the McLeod gauge. For the measurement of low pressures, the mercury was raised in limb DE to fill some or all of the bulbs, and in limb EF to one of the etched marks above the inserted join F, the compression ratio being chosen so as to give the greatest reading accuracy. In the extreme case, quantities of vapour of the order of cubic millimetres at N.T.P. could be measured with an error much below 1 %, by imprisoning them between the etched marks immediately on either side of the bent capillary at E.

Settings of the mercury to the etched marks on the gauge tube were invariably made with a rising meniscus, and readings were rejected if the difference between the meniscus heights exceeded 0·05 mm Capillarity corrections were applied throughout. It is probable that errors exceeding 0·03 mm. in the direct pressure readings were thus avoided.

Calibration, and control of errors. The volumes enclosed between the various etched marks (mark 6 excepted) were calibrated with mercury, due regard being paid to the direction of curvature, under working conditions, of the mercury meniscus set to each The adsorption vessels were calibrated by measuring out quantities of hydrogen in the burette, and then determining its pressure after distribution via the inserted join F. To control the errors involved in this indirect calibration, a blind tube of about the same capacity as the adsorption vessels was calibrated with mercury, sealed to the apparatus, and recalibrated by a hydrogen distribution experiment. The mean of six concordant determinations gave a result differing from the mercury calibration by no more than 0·080 c c. Though the possible adsorption of a little hydrogen would lead to larger errors than this in the case of the mica tubes, their effect in terms of quantities of vapour measured at relatively low pressures would again be very small.

Several experiments were made, more particularly with benzene vapour, to estimate the errors involved in the use of the McLeod principle in measuring the pressures indirectly, and in calculating quantities of vapour on the basis of the gas laws. With benzene these tests showed Boyle's law to be valid within about 0·5 % up to 95 % of the saturation pressure; thereafter systematic deviations occur, probably arising from a wall effect. With methyl alcohol the errors involved are rather greater,* but they are quite insufficient to invalidate the conclusions to be drawn from the adsorption measurements with this substance.

In accordance with the findings of Coolidge (1924), who used a quartz suspension gauge to control the errors involved in measuring low pressures of vapours by the McLeod principle, observations relating to pressures $< 0{\cdot}02$ mm. are omitted from this paper, even though they are supported by the data for slightly higher pressures. In the range of moderate pressures the direct readings provided the necessary check on the indirect measurements, and showed these to be free from systematic error, provided a suitable compression ratio was chosen.

---

* Cf. Bangham, Fakhoury and Mohamed (1934), where data are given for a rather similar apparatus.

The mica. A supply of clear muscovite mica of about 0·05 mm. thickness was obtained from a firm which had undertaken that the sheets should not be fingered either during splitting or packing. The sheets were cut with cleaned scissors and forceps into strips about $1{\cdot}5 \times 18$ cm, and tied with wire into bundles, care being taken to avoid contact with the fingers. Strips showing jagged edges or interference colours were rejected. Three adsorption vessels, A, B and C, were used at different stages of the work.

Vessel A,	area of mica	$1{\cdot}8 \times 10^4$ cm.^{2};	dead-space volume	68·1 c.c.
,, B,	,,	$1{\cdot}9 \times 10^4$ cm.^{2};	,,	89·7 c.c.
,, C,	,,	$2{\cdot}4 \times 10^4$ cm.^{2}.		

In vessel A the strips were packed tightly together in bundles with fine copper wire; 1·8 g. of copper was used, its superficial area being 25 cm.^{2}. The vessel itself had a wall area of 220 cm.^{2}.

In tubes B and C the mica strips were separated by fine wires. From wire of *ca.* 0·1 mm. diameter three spirals were first prepared, each containing as many turns as there were strips of mica to be packed in the bundle. With the aid of pins and a wooden frame, these spirals were held in slight tension (their axes being parallel), and the strips were inserted between the turns. On releasing the pins the strips became firmly gripped between the coils, and it was then a comparatively easy matter to secure the bundle with wires passed round it opposite the two end spirals. On viewing the bundles from the sides, no contacts between the strips could be observed. For tube B 6·7 g. of copper wire sufficed, its superficial area being *ca.* 300 cm.^{2}

To degas the mica, the tubes, after being sealed to the apparatus, were heated to 280° C. and pumped for several days; in between whiles they were flushed out with the experimental vapour. The pressure, as shown by the McLeod gauge, became unreadably small at the end of this period.

The experimental vapours. The choice of working substances was dictated by the desirability of knowing something of the molecular cross-sections in different possible orientations. Benzene, methyl alcohol, and carbon tetrachloride were chosen. All were rigorously purified and dried before use, and repeatedly fractionated after sealing in the apparatus until quite free from dissolved gases.

4. EXPERIMENTAL RESULTS· STRUCTURE OF THE MONOLAYERS

The observed adsorption is conveniently expressed in cubic millimetres of vapour at N.T.P., and the symbol q will be used to denote its value in terms of this unit. It will be noted that q depends, *ceteris paribus* on which of the three adsorption vessels the measurement refers to, for the area of the mica was different in each. The quotient of the adsorbing area by the number of molecules adsorbed, that is, the area per molecule, is measured in Angstrom units and denoted by A. Pressures of vapour are in millimetres of mercury and denoted by p.

The analysis of the results is complicated by the appearance of break-points in the isotherms and the formation of multimolecular films, phenomena which are discussed in a later paper. Nevertheless the experiments yielded two sets of data, the portions of the isotherms of benzene and methyl alcohol in the pressure range $< 0{\cdot}04$ of saturation, which repay comparison with theory; carbon tetrachloride, even in this low pressure range, gave quite anomalous results.

Comparison with theoretical isotherms. The theoretical isotherms to which we shall refer the data are: (1) the Langmuir equation

$$p = \text{constant} \times \frac{q}{q_\infty - q}, \tag{1}$$

where q_∞, the limiting adsorption at very high pressures, is identified by Langmuir as the number of adsorbing centres or "elementary spaces" measured in the same units as q; and (2) the equation

$$\log_e p = \log_e \frac{q}{q_\infty - q} + \frac{q}{q_\infty - q} + \text{constant}$$

or

$$p = \text{constant} \times \frac{q}{q_\infty - q} \exp\left[\frac{q}{q_\infty - q}\right] \tag{2}$$

which applies to a film of *mobile* adsorbed molecules between which only short-ranged "collision" forces are acting (Bangham and Fakhoury 1931).*

* The isotherm applies to a surface phase obeying the equation of state

$$F(A - B) = RT,$$

where F is the two-dimensional pressure; a constant differential heat of adsorption is implied.

Here q_∞, the limiting adsorption at high pressures, is unrelated to the lattice constants of the adsorbent, but the corresponding value of A, which we shall call A_∞, is the "incompressible" cross-sectional area of the adsorbed molecules, and should be slightly less than the value calculated, for example, from X-ray measurements.

For the purposes of comparison it is convenient to plot the data on a double logarithmic scale, and then to superpose the theoretical curves, similarly plotted on transparent paper, so as to obtain the best fit. The graphs of fig. 3 (benzene) and fig. 4 (methyl alcohol) have been obtained in this way, $\log_{10} p$ being the abscissa and $\log_{10} q$ the ordinate.

It will bear emphasis that the graphs of equations (1) and (2) are so very similar that if both the constants in each are treated as adjustable parameters, any set of data, unless of very wide range, must necessarily agree tolerably well with the one if it does so with the other.* The values of q_∞ obtained with the aid of the equations, however, are markedly different. only by finding out whether these q_∞ values are related to the lattice constants of the solid or to the dimensions of the adsorbed molecules can one decide whether the latter are fixed or mobile.

Structure of the mica surface. The crystal structure of muscovite has been examined by Jackson and West (1930). In the uncleaved crystal each potassium is surrounded by twelve oxygens arranged in two hexagonal rings, one on each side of the plane where cleavage is to take place, these oxygens form part of the bases of the silicate tetrahedra, to the hexagonal network of which the mica owes its characteristic structure.

In the sheets of potassium atoms the distance between each K and the next is 5·18 A, so that, the arrangement being a hexagonal one, the area per K becomes $5 \cdot 18^2 \sqrt{3}/2 \approx 23$ A^2. It is to be expected that when cleavage takes place these K atoms distribute themselves fairly evenly between the opposite surfaces, so that the area to be assigned to each is then 46 A^2. The radius of this atom being 2·31 A, and that of the K$^+$ ion only 1·33 A, it follows that the potassiums will be far from covering the surface completely: the remainder will be occupied partly by close-packed oxygens forming the bases of the silicate tetrahedra, and partly by "holes" from which potassiums have been removed; at the foot of each "hole" is a hydroxyl.

Since the potassiums are at least partly ionized,† it is to be expected that strong electrostatic forces will subsist at the cleavages, and that their polarizing effect will reinforce the "dispersion" forces in producing the adsorptive field. With methyl alcohol there is also some possibility of hydroxyl bond formation, either with the partly exposed hydroxyls of the lattice or with chemisorbed water molecules. The part played by the latter remains an unknown factor.

---

* The effect of the exponential term of equation (2) on the *shape* of the double logarithmic graph is only to render it rather less concave to the log p axis near the point of maximum curvature.

† Jackson and West (1930) (who, however, report the measurement as open to some doubt) found 17·7 electrons per K in the uncleaved crystal.

Low-pressure isotherms of benzene. In fig. 3 there are graphed the experimental data for benzene at 25 and 35° C, together with the following theoretically derived curves:

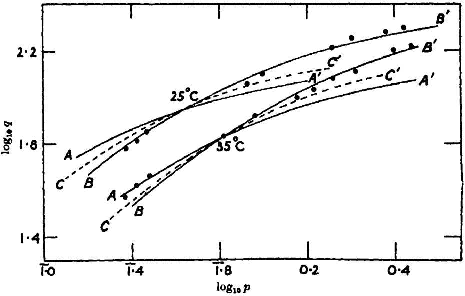


Fig. 3. Isothermals of benzene on mica. The broken lines CC' are graphs of Langmuir's isotherm, allowing 46 A² per elementary space Curves AA' are for a film of mobile molecules lying flat ($A_\infty = 36$ A²); curves BB' for mobile molecules on edge ($A_\infty = 21{\cdot}5$ A²).

(1) The broken curves CC' which are graphs of Langmuir's equation if 46 A² is allowed for each elementary space. The agreement obtained here is poor; it would be even worse if only 23 A² were allowed per elementary space.

(2) The full lines AA' derived from equation (2) for a film of mobile molecules of incompressible area 36 A², which is approximately the area to be assigned to the benzene molecule if lying flat.

(3) The full lines BB' which are graphs of the same equation, but for molecules of incompressible area 21·5 A², which is the cross-section of the benzene molecule perpendicular to the plane of the ring (Cox 1932; see

Adam 1930). Here the agreement with experiment is tolerably good. It is even better with $A = 20$ A^2, which is quite a reasonable estimate since the limiting adsorption values give the molecular areas under high compression.

We conclude that the molecules forming the monolayer of benzene are freely mobile (as indeed one would expect them to be on chemical grounds, compound formation being highly improbable), and that over the range of concentrations examined they are oriented on edge; it is not unlikely that at still lower concentrations they tend to lie flat.*

Objection may be raised to the above conclusions on the grounds that the effects of intermolecular forces have been ignored, and that the deviation from the Langmuir graph may be due to the formation of a second molecular layer. Neither objection is as serious as appears at first sight. Evidence will be given in a later paper that the formation of polymolecular films sets in with something of the abruptness of a phase transition; the differential heat of adsorption, *which is very nearly constant* over most, if not all, of the range here considered, then falls sharply to a value very near the normal heat of condensation.

Low-pressure isotherms of methyl alcohol. The experimental isotherms for methyl alcohol graphed in fig. 4 lead to very similar conclusions. To obtain agreement with the Langmuir equation, which when plotted on the double logarithmic scale gives the broken curves in the figure, it is necessary to suppose that each elementary space occupies 17·7 A^2, a figure which bears no obvious relation to the constants of the lattice. The equation for a mobile film (full line) requires $A = 11{\cdot}4$ A^2. Bearing in mind that the use of this equation necessarily leads to a low estimate of the molecular cross-section, this is a not unreasonable figure for the end-on orientation of the molecules, the closeness of packing of which would then be determined by the cross-sectional area of the CH_3 groups.† Using the data of Langmuir (1918), Bangham and Fakhoury (1931) found 14 A^2 for the cross-sectional area of the CH_4 molecule by applying the same method.

---

* The graph of equation (2) for $q_\infty = 355$ mm.^{3}, $A_\infty = 20$ A^2, which has been omitted from fig. 3 for the sake of clarity, gives almost perfect agreement with observation except for the three lowest points on the 35° isotherm, which fall rather high. The measurements of the adsorption energy referred to in the text did not extend to quite such small concentrations, so that a different orientation in the region of very small covering is not precluded

† Hendricks (1930) has shown that in a whole series of *solid* compounds the distance of closest approach of two carbon atoms belonging to different molecules lies between 3·6 and 3·9 A. In a planar arrangement of one-carbon molecules in hexagonal close-packed array the area to be assigned to each would therefore be between 11·3 and 13·2 A^2.

Though possibly not quite conclusive, these measurements strongly suggest that the methyl alcohol molecules are also freely mobile in the monolayer, and that the statistically favoured orientation is that in which the CH_3 groups are pointing outwards, and the hydroxyls turned in towards the surface. This orientation was *not* found with methyl alcohol adsorbed on *charcoal*, where, both in the "primary" films formed at low concentration, and in the denser ones at higher pressures, the evidence showed the axes of the molecules (not the dipole axes) to lie parallel with the surface (Bangham 1934). The different orientation at the surface of mica probably accounts for the surprising absence of any evidence of association in the adsorbed phase, at all events below the first breakpoint. Both on charcoal and on mercury the molecules of all the lower alcohols appear to be associated to a considerable extent, even at low concentrations. With mica this is certainly not the case,* and the tendency of the molecules to

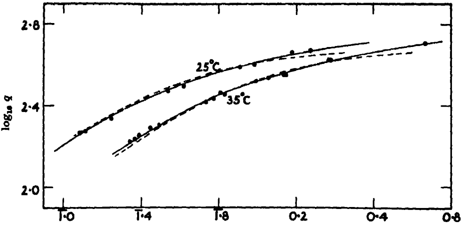


Fig. 4. Isothermals of methyl alcohol. The broken lines are graphs of Langmuir's isotherm, allowing 17·7 A² per elementary space; the full lines are for film of mobile molecules oriented end-on ($A_\infty = 11{\cdot}4$ A²).

---

* To obtain further evidence on this point, some adsorption measurements were made at pressures well below the range to which fig. 4 refers. According to Coolidge's already quoted statement as to the limits of accuracy of the McLeod gauge, these should be liable to a considerable systematic error, and they are not reproduced here. Use was made thereof, however, to obtain values of the surface energy lowering F with the aid of the integrated form of the Gibbs adsorption equation discussed recently by one of us (Bangham 1937); here even a considerable systematic error at small adsorptions would lead to no very serious consequences in the range of denser covering. The graphs of FA against F were then plotted. These were found to be nearly linear, and to show no trace of the initial downward bend so characteristic of similar graphs for the alcohols adsorbed on charcoal. Such as it is, the evidence would suggest that Henry's law is obeyed by methyl alcohol on mica at pressures ~ 10^{-3} mm.

cluster together, which probably accounts for the peculiarities of the carbon tetrachloride isotherms now to be described, is entirely absent with benzene and methyl alcohol in the range of concentrations we have been considering.

Experiments with carbon tetrachloride. The adsorption vessel B, containing separated mica strips, was used. The adsorption isotherm at 35° C. is graphed in fig. 5, the abscissa variable being the pressure p. The graph shows marked discontinuities, recalling, on an enlarged scale, those found by Allmand and his co-workers (Allmand and Chaplin 1930, Allmand, Burrage and Chaplin 1932). The steeply rising sections of the graph, which are not improbably all concave to the adsorption axis,* point to a strong

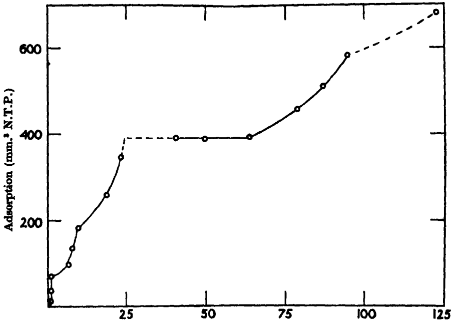


FIG. 5. Isothermal of carbon tetrachloride at 35° C.

---

* The broken line joining the points in the figure is, of course, partly hypothetical, the points being too widely spaced to permit interpolation; the form given to the isotherm was suggested by the course of the curves immediately preceding the first and fourth breakpoints, which are both markedly concave to the adsorption axis. The experiments with carbon tetrachloride were carried out as rapidly as possible so as to avoid interaction between the vapour and the copper wire used for separating the plates; for unpublished experiments by N. Fakhoury in these laboratories have shown that carbon tetrachloride is liable to slow decomposition at certain metal

tendency for the molecules to cluster together under the influence of intermolecular attractive forces. The breaks occur when the areas per molecule lie between the following limits:

First,	$A = 72\text{–}101$ A^2,	Third,	$A = 17{\cdot}8\text{–}20{\cdot}4$ A^2,
Second,	27–39,	Fourth,	12·2–10·4.

In normal liquid carbon tetrachloride at 35° C. each molecule occupies about 163 A^3 (molecular volume divided by Avogadro constant), and they probably group themselves into a more or less close-packed arrangement (Menke 1932). It will be seen that the first break occurs when the first molecular layer is by no means complete If we suppose, following Langmuir, that the mica presents more than one kind of "elementary space", so that its surface is heterogeneous when viewed according to the scale represented by the molecule of carbon tetrachloride, it becomes an easy matter to account for the breaks, but the marked grouping tendency of the molecules presents a difficulty. Moreover, the areas per molecule at the breakpoints bear no obvious relation to the lattice constants of the mica.

If the molecules forming the first layer have no definite points of attachment, it is probable that the statistically favoured orientation is one in which each CCl_4 has three Cl atoms in contact with the surface and with the Cl atoms of its neighbours. Given that as the concentration rises the molecules tend to group themselves into a close-packed arrangement of tetrahedra, it is quite comprehensible that discontinuities should arise, though the occurrence of the first of these at such a low concentration again presents a difficulty.

If we assign to the adsorbed phase a density equal to that of the bulk liquid, the thickness of the film would lie between 4·2 and 6·0 A at the second break, between 8·0 and 9·2 A at the third, and between 13·4 and 15·7 A at the fourth. These are not widely different from the thicknesses to be expected if each breakpoint coincided with the completion of a molecular layer; but it will again bear emphasis that any attempt to build up a model of the adsorbed phases on the same lines as a liquid is necessarily incomplete unless the theory takes into account the fact that the film is incapable of acting as a nucleus of condensation of bulk liquid phase.

surfaces at comparatively low temperatures. In spite of the greater rapidity of working, we have no reason to question the accuracy of the data, which were obtained when the technique of measurement was well established, and is well supported by the appearance of certain regularities, to be discussed elsewhere; the latter also indicate that no appreciable interaction with the copper can have taken place.

One of the authors is grateful to the Leverhulme Trust Committee for the award of a Fellowship, which has expedited the long-delayed publication of these results of experiments which were carried out in the laboratories of the Egyptian University during the years 1931–7.

5. SUMMARY

Measurements have been made of the quantities of benzene, carbon tetrachloride, and methyl alcohol adsorbed at a known surface area of mica at pressures ranging from 0·02 mm. to near saturation. The general shape of the benzene isotherm is sigmoid, convex to the adsorption axis at lower pressures, where the monolayer is incomplete, but becoming markedly convex to this axis near saturation, where the film is polymolecular. The same isothermal is obtained when the mica plates are separated by fine wires as when they are tightly packed together. Capillary condensation in incipient cleavages at the mica edges does not take place, and the polymolecular films have properties which differentiate them from the bulk liquids. The isothermals of benzene and methyl alcohol at lower pressures agree well with the theoretical equation for films of mobile molecules oriented end-on to the surface, but Langmuir's equation leads to limiting adsorption values which bear no relation to the lattice constants of the mica (muscovite). The isothermals of carbon tetrachloride shows marked discontinuities, as also do those of benzene and methyl alcohol over the range of transition to polymolecular films.

REFERENCES

Adam 1930 "The Physics and Chemistry of Surfaces", p 96. Oxford.
Allmand, Burrage and Chaplin 1932 *Trans. Faraday Soc.* **28**, 218.
Allmand and Chaplin 1930 *Proc. Roy. Soc.* A, **129**, 257.
Bangham 1934 *Proc. Roy. Soc.* A, **147**, 175.
— 1937 *Trans. Faraday Soc.* **33**, 805.
Bangham and Fakhoury 1931 *J. Chem. Soc.* p. 1324.
Bangham, Fakhoury and Mohamed 1934 *Proc. Roy. Soc.* A, **147**, 164.
Bawn 1932 *J. Amer. Chem. Soc.* **54**, 72.
Bragg, W. L. 1937 "Atomic Structure of Minerals." Oxford.
Coolidge, A. S. 1924 *J. Amer. Chem. Soc.* **46**, 596.
Cox, E. G. 1932 *Proc. Roy. Soc.* A, **135**, 491.
Hendricks 1930 *Chem. Rev.* **7**, 431.
Hückel, E. 1932 *Trans. Faraday Soc.* **28**, 382.
Jackson and West 1930 *Z. Kristallogr.* **76**, 211.
Joffé 1928 "The Physics of Crystals." New York.
Langmuir 1918 *J. Amer. Chem. Soc.* **40**, 1361.

Lennard-Jones and Dent 1928 *Proc. Roy. Soc.* A, **121**, 247.
McBain 1932 "The Sorption of Gases by Solids", p. 329.
Menke 1932 *Phys. Z.* **33**, 593.
Obreimoff, J. V. 1930 *Proc. Roy. Soc.* A, **127**, 290.
Polanyi, M. 1933 *Phys. Z. Sowjet.* **4**, 144.
Smekal 1925 *Phys. Z.* **26**, 707.

The influence of rate of deformation on the tensile test with special reference to the yield point in iron and steel

By C. F. Elam

(*Communicated by Sir Harold Carpenter, F.R.S.—Received 26 January 1938*)

Part I

The following experiments were carried out with two principal objects in view: (1) to investigate the deformation of those metals, particularly iron and steel, in which the stress-strain curve does not immediately rise at the onset of plastic distortion; (2) to determine the effect of rate of deformation on the yield and subsequent stress-strain curve.

It is impossible to give an adequate summary of the literature which deals with this subject, but a bibliography is included in an appendix and some of the most important results are referred to briefly below.

The first description of a fall in the load at the yield point in iron and steel was made by Bach (1905). Since that time a similar effect has been observed in zinc and cadmium crystals (Becker and Orowan 1932; Orowan 1934; Schmid and Valouch 1932), copper and copper alloy (Elam 1927) crystals, sodium chloride (Classen-Nekludowa 1929; Dawidenkow 1930; Joffé 1928), α brass (Elam 1927; Sachs and Shoji 1921), both single crystals and polycrystalline material, and duralumin (Dawidenkow 1930). The conditions affecting the occurrence of an upper yield point have been most fully investigated in the case of iron and steel, and of these the most important are: rate of application of load (Körber and Pomp

1934; Kuhnel 1928; Moser 1928; Quinney 1934, 1936; Siebel and Pomp 1928); shape of test-piece (Bach 1904; Docherty and Thorne 1931; Kuhnel 1928; Quinney 1934, 1936); axial loading (Docherty and Thorne 1931; Quinney 1934, 1936); and heat treatment of the metal (Edwards and Pfeil 1925; Emsslin 1928; Kuhnel 1928). The value of the upper yield point may, under favourable conditions, be raised momentarily above that of the breaking stress under a static test (Ginns 1937; Hopkinson 1905; Quinney 1934, 1936). The value of the lower yield is less affected by the conditions enumerated above and remains approximately constant for the material (Emsslin 1928; Kuhnel 1928; Moser 1928). It is not always possible to distinguish an upper and lower yield on certain testing machines, and this fact has recently led to much discussion, particularly in Germany, on the relation between testing machine and results obtained (Bernhardt 1936; Pomp and Krisch 1937; Siebel and Schwaigerer 1937; Späth 1937*a*, *b*; Welter 1935). Welter (1935) started the controversy by stating that the type of curve obtained depended on the elasticity of the machine. He and other workers compared the results of tests by direct loading, in beam type machines and in those of the type originally devised by Polanyi (1925) and elaborated by Quinney (1934, 1936), in which the load is applied by pulling against a spring. Tests were also carried out with a spring placed between the shackles of a lever testing machine together with the test-piece (Siebel and Schwaigerer 1937; Späth 1937*a*, *b*; Welter 1935). This method may be compared with that of Robertson and Cook (1913; Cook 1931) who made use of a weigh-bar in their measurements. Different results were obtained with the different methods employed. These experiments also demonstrated the importance of rate of deformation in connexion with the yield point where a large increase in rate of flow takes place (Bernhardt 1936; Körber and Pomp 1934; Kuhnel 1928; Pomp and Krisch 1937; Siebel and Schwaigerer 1937).

The effect of duration of test has been studied chiefly in connexion with creep phenomena and the rate of flow at temperatures higher than normal or on metals with low melting-points. Experiments at constant stress led Andrade (1910, 1914) to put forward an empirical formula connecting change of length of a test-piece with time. The effect of velocity of deformation on the stress has shown that the stress is greater at higher speeds (Deutler 1932; Ludwik 1909). A change from a fast to a slow rate of extension has been compared with the afterworking of metals (Ludwik 1909), and of amorphous substances (Braunbek 1929), which are very susceptible to velocity of deformation.

In the Becker-Orowan (Orowan 1934, 1935*a*, *b*, 1936) theory of hardening the rate of deformation is considered to vary with the externally applied stress and this in turn determines the shape of the stress-strain curve.

A stress-strain curve in which the load temporarily falls, or does not immediately rise with deformation, is very similar to the deformation by jumps (Becker and Orowan 1932; Classen-Nekludowa 1929; Dawidenkow 1930, Joffé 1928; Orowan 1934, 1935*a*, *b*; Schmid and Valouch 1932) which is characteristic of certain crystals and polycrystalline materials. The magnitude and frequency of the jumps are remarkably uniform and they may be preceded by a building up of the stress caused by some hindrance to deformation, just as the yield point in iron is exceeded in certain circumstances. The extension associated with each jump varies from 5 to 200μ.

Another feature in common is the localized nature of the deformation which is associated with the formation of Lüder's Lines in iron and steel (Docherty and Thorne 1931; Fell 1935; Kuntze and Sachs 1928; Nadai 1931).

PART II. INTRODUCTION

Tensile testing machines

Two types of testing machine were used; a 50 ton Buckton and two autographic recording machines, referred to as 5 and 6 ton machines, made in the Engineering Department to the design of Mr Quinney. The latter have been fully described elsewhere (Quinney 1934, 1936), and it is only necessary to refer to the mechanism in so far as it affects the results of the tests carried out on them.

The results of calibration of load, extension, etc., will be found in Tables II–IV.

The rates of extension in the different gears were determined by means of a stop watch and measuring microscope and an automatic counter attached to the worm gear was also calibrated, and the time of each test and readings of the counter were made throughout the experiments. The rates of extension at the higher speeds, i.e. gears No. 2, 3 and 4 on the 5 ton machine, varied slightly with load owing to the slowing up of the motor. The arrangements on the 6 ton machine made it unsuitable for tests where rates of extension were required to be measured.

The record which is obtained on the smoked glass has one ordinate curved, as the pointer measuring the load moves along an arc of a circle of 35 in. radius. It must be measured in a similarly constructed machine which has also been calibrated with the testing machine. One of the diagrams has been superimposed on a chart showing the relation between the record and the load and extension (fig. 2).

The chief error was found to arise in the strain measurements due to sticking of the box holding the plate, tapes, etc., and to the fact that the bedding down of the heads of the test-piece and the differential extension of the shoulders are included in the total extension measured. Some of these difficulties cannot be overcome with the present design of the machine but the magnitude of the error can be ascertained.

In order to calibrate the extension, a travelling microscope was adjusted to read between razor blades attached to each shackle, and a direct reading of the movement of both shackles was made and compared with the movement of the plate. One set of readings was taken with no test-piece in position, in which case only one shackle moved. When a test-piece was being pulled both shackles moved; that attached to the spring giving the movement due to the deflexion of the spring, whereas the difference between readings on both razor blades gave the total extension of the test-piece. By carrying out a number of measurements of the extension both loaded and unloaded, it was found that there was always a difference in the zero when readings on loading and unloading were compared. On the other hand, extension measurements from stage to stage of a test are reasonably accurate and agree with independent measurements by means of scratches on the surface of the test-piece. In the course of a normal test, however, readings are required in both directions, particularly where an extension of the test-piece causes a return of the spring, such as occurs at the yield point of iron and steel. The total elastic extension was also found to be inaccurate and the figures given in the tables are only approximate.

As all the tests and measurements were made in the same way, the results may be compared, even if errors are included in the absolute values given in the tables.

Shape and dimensions of test-pieces

The "streamlined" specimen advocated by Docherty and Thorne (1931) and by Quinney (1934, 1936) has serious disadvantages when extension measurements are required and any accurate correlation of load and extension during the test. The only advantage seems to be that it is possible to obtain a high value for the upper yield, but as this is liable to fluctuation from a number of causes and the variations obtained even with streamlined test-pieces are so great, the disadvantages appear to outweigh the advantages.

A normal shaped test-piece was therefore used throughout, having a parallel length of 6 in., 0·30 or 0·4 in. diameter, and heads 0·75 in. diameter and 0·5 in. long, which fitted into the ball shackles in the manner described by Quinney. In addition, some measurements were made on large bars 18 in. long, 0·75 in. diameter, and 8 in. parallel with screw threaded ends, which also fitted into ball shackles for use in the 50 ton Buckton testing machine. It was therefore possible to ensure axial loading in both machines.

The test-pieces were ground to their final dimensions so that the finish was good. For the purposes of measuring change in dimensions, they were marked usually by four fine scratches at 90° along the axis of the bar, and by a series of scratches 1 or 2 cm. apart at right angles to these. All test-pieces were heat-treated after machining by annealing *in vacuo* or, in the case of the large bars, in hydrogen. Measurements of the diameters were made by means of a micrometer, between scratches by means of a travelling microscope reading to 0·001 mm.

Material

Tests were made on armco iron (not analysed) and two steels of the following composition:

	O	T
C	0·68	0·248
Si	0·10	0·101
Mn	0·30	0·016
S	0·37	0·018
P	0·02	0·857

The copper was Post Office specification, high conductivity metal but has not been analysed as the tests were only comparative.

The duralumin was kindly given by Dr Leslie Aitcheson of James Booth and Co.

A representative series of High Duty alloys was specially prepared and supplied through the kindness of Mr W. C. Devereux.

The exact composition and treatment of these alloys are of secondary importance in the present investigation, as here also the tests are of a comparative character.

Description of experiments

(1) *Comparison of tensile tests in Buckton and Autographic machines, with special reference to the nature of deformation at the yield*

The first experiments were intended to test the conditions under which a drop in load associated with the yield point could be demonstrated in the Buckton testing machine and at the same time to measure the change in dimensions. For this purpose large bars were chosen to give the best possible accuracy of measurement of diameter. For dimensions and description of preparation see p. 572. The steel used was T, 0·248 % C (see p. 572).

The load was applied in the usual way with the beam floating just above the stop. This was found to be necessary in order to interrupt the test within the yield.

After the load at which the first yield occurred was noted the specimen was removed and measured, reloaded, and the load at which it began to yield again measured. This was repeated until the load began to rise. The effect of rest during the yield for this steel was found to be negligible.

The measurements showed that during the yield the deformation was not uniform and the diameters in two directions at right angles sometimes differed as much as 0·3 %. Luder's Lines became visible on the surface, generally starting from one end, and spreading along the parallel portion.

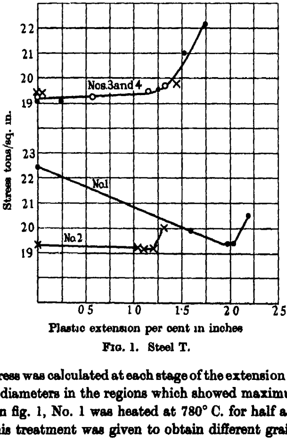


Fig. 1. Steel T.

The true stress was calculated at each stage of the extension from measurements of the diameters in the regions which showed maximum changes in dimension. In fig. 1, No. 1 was heated at 780° C. for half an hour, No. 2 at 950° C. This treatment was given to obtain different grain sizes. Tests on two other bars, both heated at 900° C., fig. 1, Nos. 3, 4, gave very consistent results both as regards value of stress at which plastic yielding continued at different stages of the yield and for length of yield which was approximately 1·25 % in the case of the bars heated at 900° C. and 1·9 % for those at 780° C.

Table I shows the spread of the extension during the yield. These measurements were made in Quinney's machine on a smaller bar (0·4 in. diameter, 6 in. parallel) and the travelling microscope was mounted so as to take readings while the specimen was still in the machine.

TABLE I. NO. 7 T, HEATED 750° C. INCREASE IN LENGTH OF SECTIONS OF 2 CM.

	1st ext.	2nd ext.	3rd ext.	4th ext.	5th ext.
1	0	0	0	0·002	0·040
2	0	0	0	0·036	0·042
3	0	0	0	0·032	0·044
4	0	0	0·023	0·034	0·042
5	0	0·017	0·032	0·036	0·046
6	0 003	0 032	0·035	0·035	0·042

The end of the yield, indicated by the diagram, came between the fourth and fifth extensions, by which time the distortion had become uniform, gaps between the Lüder's Lines gradually having been "filled in". There is every reason for assuming that the extension in each band is equal to the total yield.

These experiments confirm the work of Kuntze and Sachs (1928) on the change in dimensions which takes place through the yield. The tendency of the load to fall after the initial yield, can be shown if proper precautions are taken, and it cannot be attributed to change in cross-sectional area. Those bars in which the load did not fall, continued to deform throughout the yield at an almost constant stress, and this is not surprising since the metal in the unchanged part of the bar is still un-work-hardened. The observations of Edwards and Pfeil (1925) on the connexion between grain size and length of yield are also confirmed.

Some tests were also made with a ⅜ in. square bar of armco iron and a mild steel in order to watch more closely the formation of Lüder's Lines. These were polished on all four sides and marked in order to obtain extension figures. The angles which the lines made on the faces were also measured, and were found to vary from 40 to 88°. They were not the same on opposite faces but in some cases could be traced round the intersection of two neighbouring faces. Assuming that in such a case the marks represented the traces of a plane on the two faces, the inclination of the plane to the axis could be ascertained. The planes varied from 40 to 50° to the axis. The fact that, even in a polycrystalline metal, there is a tendency to slip along planes of maximum shear stress, is in agreement with the observation of the formation of an ellipse in the round bars noted above. In most cases two sets of bands were formed, but this would also cause a thinning of the bar in one dimension if the deformation were confined to two planes at 180° to each other, i.e. on opposite sides of the bar. The amount of deformation that occurs in this manner must be very small.

Pyramid hardness tests were carried out on these bars in the region showing Lüder's Lines. In the case of armco iron the extension at the yield which was approximately 2%, gave an increase of 11% (Fell 1927, 1935). In a mild steel an increase of 7·5% was obtained with a yield of 2·4%. In both cases the measurements were made about 2 weeks after straining.

(2) *The effect of rate of deformation on the stress-strain curve (iron, steel, copper)*

Tests at different speeds were carried out on armco iron, two steels (O and T) and copper. The results are given in Tables II–IV and selected diagrams in figs. 2–6. These are obtained by printing through two films at once.

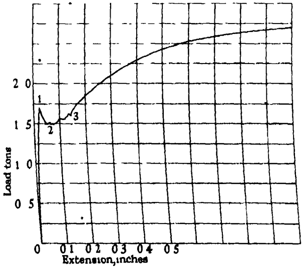


FIG. 2. Steel O. 7 = gear R.R.

Great care was taken to compare test-pieces which had been cut from the same bar and heat-treated simultaneously. Considerable variations occurred in spite of all precautions, showing how sensitive the deformation must be to slight differences in individual tests.

Full details of the measurements made in the case of iron and steel are given in explanation of the tables, but it may be desirable to state briefly

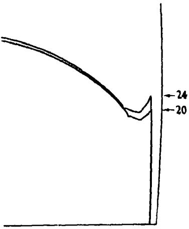


FIG. 3. Armco iron.
20 = gear R.R.; 24 = gear 3.

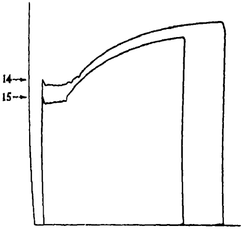


FIG. 4. Steel T.
14 = gear 4; 15 = very slow.

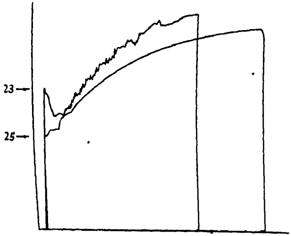


FIG. 5. Armco iron. 23 = gear 4; 25 = very slow.

why these values were chosen for comparison. Previous work has shown that an increased rate of loading raises the upper yield point but the relation between this point and the subsequent extension is not so well known. The first measurement required was the highest point reached at the yield

(1). Secondly, the lowest load shown on the diagram after yielding had started (2), and the point at which this occurred in the yield, which is indicated by the extension (1–2). It has been pointed out already on p. 571 that an extension of the test-piece results immediately in a return of the spring and a drop in load on the record but that exact numerical agreement between the two values was never obtained owing to errors in the machine. Moreover, while a test is in progress the end of the test-piece attached to the spring is being moved in the opposite direction and the record merely indicates the result of these combined movements. It is not surprising therefore to find that the average value for the lower yield also varies with the rate of testing and also that the lowest point may occur anywhere within the yield although in most instances it is nearer the upper yield. That is to say, maximum extension usually follows immediately the onset of plastic distortion, and is greater the higher the initial yield point. The ups and downs within the yield are associated with the propagation of the distortion by the formation of Lüder's Lines. The value of the load at the end of the yield shows if the metal has hardened during this stage of the deformation, and the figures in the column headed 1–3 give the length of the yield. These stages cannot be separated in armco iron and are entirely absent in the tests at very slow speeds in both armco iron and the steel O. They persist, however, in steel T even at the slow speed of testing.

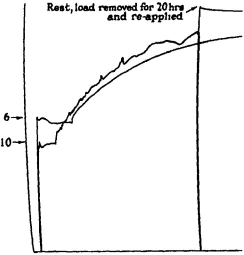


FIG. 6. Steel O. 6 = gear 4; 10 = very slow.

The results contained in Tables II–IV and illustrated in figs. 3–6 may be summarized as follows:

(1) The value of the upper yield point is raised by increased rates of deformation.

(2) The value of the lower yield is raised by increased rates of deformation.

(3) The length of the yield is increased by increased rates of deformation.

(4) The horizontal portion of the stress-strain curve may disappear in armco iron and a very low carbon steel if the rate of deformation is sufficiently slow.

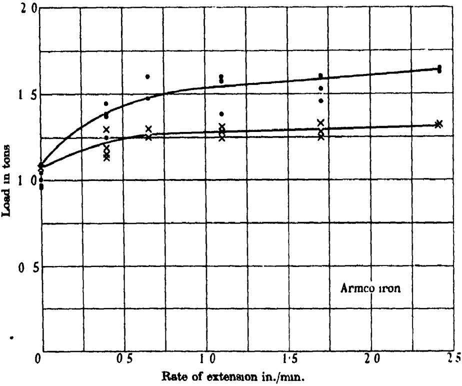


Fig. 7

In fig. 7 the values of upper and lower yields are plotted against rates of deformation for armco iron. The maximum difference occurs between the slowest rate (which is very nearly zero on the scale of the diagrams) and the slowest rate obtainable on the Wilson automatic gear-box (R.R.). Rates of deformation between these have not yet been investigated.

Apart from the presence or absence of a well-marked yield, other differences are manifest between the very slow and more normal rates of testing. The curves obtained at the slowest speed rise steeply from the elastic limit in a series of irregular jerks, each one of which consists of an excessive rise and fall in load like miniature yields, but the curve continues to rise in between. The increase in resistance to deformation is also greater for the slow rate of extension. Fig. 9 shows the results obtained from two examples of armco iron. A quick method of comparing the curves direct from the records was to print on to films and superimpose these, making a small allowance for differences in diameter where necessary. This obviated the laborious task of measuring up all the curves and plotting in the form of load-extension diagrams similar to fig. 9.

A comparison of all the diagrams by this method showed that relatively large increases in rate of hardening occurred in armco iron and the steel O (not steel T) between the R.R. gear and very slow rates of extension, but that at greater speeds all the steels showed a change in the opposite direction as the speed was increased. Moreover, a change in the middle of the test from gear No. 4 to gear R.R. caused a definite drop in stress.

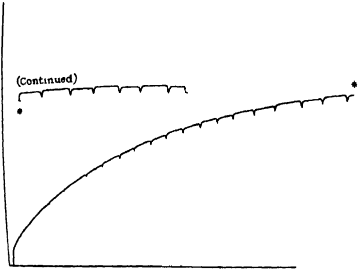


Fig. 8. Copper pulled in alternate gears R.R. and 3. Rests 0·5 min.

Similar tests were then carried out with copper and here also a change of speed altered the stress. Fig. 8 was the curve obtained when copper was extended in stages of alternate speeds, R.R. and gear No. 3, with intervals of rest of half a minute. A sudden drop in load was always observed when any test (both iron and copper) was interrupted, the magnitude of the fall depending on the rate of deformation and the stress. There was no such drop at the end of very slow tests although creep would take place in time. A very slow test on copper showed that, unlike iron, the rate of hardening was less than at high speeds (gear No. 4). The two curves are given and compared with iron in fig. 9.

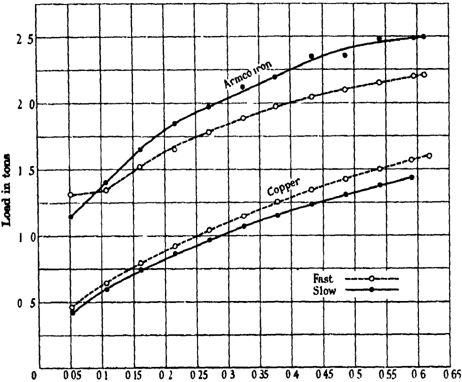


Fig. 9

The effect of rate of deformation on the stress-strain curve just described is similar to that already investigated by Ludwik (1909) and Deutler (1932), and can be applied to explain also the effect of speed on the lower yield. But the increase in resistance to deformation which occurs in armco iron and mild steel when extension is very slow presents a different problem and a solution may be sought in another well-known characteristic of iron and steel, namely, the property of hardening on resting after straining, originally investigated by Muir (1900*a*, *b*).

The effect of rest. The following experiments were carried out to test the influence of rests of varying periods at different stages of extension. Full details are given in Tables V, VI and a representative record is shown in figs. 10, 11.

It will be seen that a rest, even under load,* is followed by a new high yield which may exceed in magnitude the primary yield. This may be due partly to the better alignment of the test-piece and the smoothing out of inequalities during plastic distortion. The peak of the yield is followed by a drop back on to a curve which may be a continuation of the previous portion, or there may be a definite step up, according to the material and the length of rest. There was always a "step up" with armco iron, but

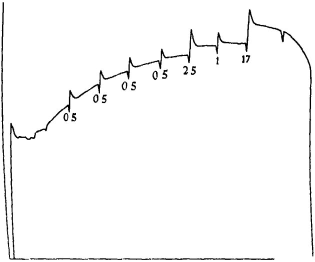


FIG. 10. Steel No 11 O Gear R.R. with rests Figures refer to duration of rests in hours.

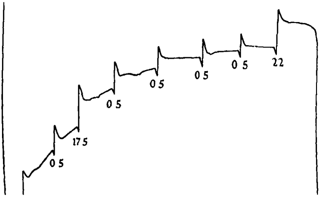


FIG. 11. Armco iron No. 12_2. Gear R.R. Figures refer to duration of rests in hours. (The bottom part of diagram has been cut off to reduce size.)

---

* Experiment showed that the same result was obtained if the load was removed during the test.

only after long rests with the steels. In fact, the curve with short rests was identical with a normal curve if the rates of extension were similar. There seems to be a tendency for the increase of hardness to become less for the same period of rest at the end of the test, but it is impossible to estimate the effect of previous rests on the subsequent behaviour of the specimen. The rate of hardening falls off with time, but it is quite possible that if a sufficiently long rest were given at each pause to enable the maximum hardness to develop the increase in hardness would be closely related to the amount of extension. This point cannot be decided from the present experiments as extensions between rests were not always the same.* A large number of separate tests on a series of specimens is required before the factors contributing to the increase in hardness are isolated and this involves introducing yet another variable in the use of different test-pieces. For the present, it may be said to be sufficient to have established the fact that the iron and steel used in these experiments, hardens on resting, to an amount closely related to the length of pause, and that these results are in agreement with the effect of a very slow rate of deformation in producing an increased rate of hardening.

Similar tests with copper showed that this metal did not harden on resting and confirmed the observation that the rate of hardening at slow was less than at fast speeds.

Tables II–VI

Tensile tests carried out on autographic machine (5 tons).

Calibration of Machine.

Extension:
1 in. plate = 0·216 in. extension between shackles.
Load·
1 in. plate = 0·4675 ton.
Deflexion of spring = 0·0675 in./ton.
Rates of extension:†
Very slow = 0·000204 in./min.
Gear R.R. = 0·0405 in./min.
Gear 1 = 0·0661 in./min.
2 = 0·110 in./min. } Wilson automatic gear-box.
3 = 0·170 in./min.
4 = 0·242 in./min.

---

* There was no means of ensuring equal extensions except by timing and this was difficult over short periods. It was successful in No. 12 O, when the periods of extension were 40 sec.

† At rates of extension higher than gear No. 1 the load affected the values obtained. Those given are average for the tests made.

The numbers in headings refer to points measured on the curves as follows (see fig. 2):

0 = beginning of record.

1 = top of peak and beginning of plastic extension.

2 = lowest value of load given by pointer after plastic extension has begun. This is usually close to 1, but may occur at any point in the yield

3 = end of yield, i.e. point at which curve begins to rise finally.

Size of test-pieces, 0·30 or 0·40 in. diameter, 6 in. between shoulders.

TABLE II. ARMCO IRON, HEATED *IN VACUO* AT 750° C. FOR 1 HR. 3–7 AND 8–11 HEATED TOGETHER. NO DEFINITE END OF YIELD OR HORIZONTAL PART OF CURVE

			Load (tons)		Extension (in.)	
No.	Diam. in.	Gear	1	2	0–1	1–2
7	0·400	Very slow	0·977		– –	– –
8	0·401	,,	0·966		–	––
9	0·401	,,	1·04	1·005	0·0094	–––
11	0·400	,,	1·01		0·0126	–
5	0·399	R.R.	1·25	1·18	0·0165	0·0425
4	0·401	2	1·38	1·24	0·0141	0·0374
10	0·400	2	1·385	1·315	0·0155	0·0522
6	0·399	3	1·46	1·25	0·0174	0·0315

TABLE II. ARMCO IRON, HEATED *IN VACUO* AT 750° C. FOR 1 HR.

			Load (tons)			Extension (in.)		
No.	Diam. in.	Gear	1	2	3	0–1	1–2	1–3*
14	0·397	R.R.	1·38	1·15	1·31	0·0145	0·0363	0·0940
15	0·398	R.R.	1·44	1·135	1·295	0·0164	0·0294	0·0948
16	0·399	1	1·475	1·25	1·35	0·0162	0·0320	0·1053
18	0·398	2	1·575	1·31	1·43	0·0158	0·0400	0·114
19	0·397	3	1·61	1·285	1·41	0·0162	0·0354	0·109
17	0·399	4	1·635	1·32	1·425	0·0179	0·0363	0·1185
25	0·398	Very slow	1·099	1·075	1·163	0·0138	0·0048	0·0320
20	0·397	R.R.	1·37	1·29	1·24	0·0166	0·0439	0·0863
21	0·398	1	1·51	1·25	1·29	0·0138	0·0335	0·0911
22	0·399	2	1·59	1·28	1·34	0·0175	0·0400	0·0916
24	0·398	3	1·53	1·33	1·38	0·0196	0·0480	0·1055
23	0·391	4	1·64	1·32	1·36	0·0160	0·0343	0·0970

* No horizontal part of curve in armco iron, but all these test-pieces gave a slight break and change of direction of curve and this was taken as end of yield.

TABLE III. STEEL O

No.	Diam. in.	Gear	Load (tons) 1	Load (tons) 2	Load (tons) 3	Extension (in.) 0–1	Extension (in.) 1–2	Extension (in.) 1–3
Heated at 750° C. *in vacuo* for 0·5 hr.								
1	0·401	R.R.	1·585	1·53	1·55	0·0207	0 0363	0·118
2	0·399	1	1·64	1·50	1·50	0·0192	0·116	0·116
3	0 399	2	1·565	1·53	1·58	0·0177	0·103	0·130
4	0·398	3	1·73	1·565	1·63	0·0168	0·039	0·135
6	0·399	4	1·65	1 565	1·57	0·0181	0 1295	0·1315
Heated at 800° C. *in vacuo* for 0 5 hr.								
			1st 2nd					
10	0 400	Very slow	1 26 1·33	1·295	1·295	0·0199	0·0086	0·0655
7	0·400	R.R.	1·699	1·46	1·56	0·0236	0·0469	0·119
11	0 399	R.R.	1·685	1·47	1·58	0·0188	0·0660	0·1315
9	0·399	2	1·605	1·425	1·53	0·0201	0·0274	0 115
12	0 400	3	1·52	1·48	1 65	0·0173	0·0149	0·143
8	0 401	4	1 525	1·515	1·62	0·0186	0 0041	0·136

TABLE IV. STEEL T

No.	Diam. in.	Gear	Load (tons) 1	Load (tons) 2	Load (tons) 3	Extension (in.) 0–1	Extension (in.) 1–2	Extension (in.) 1–3
Heated *in vacuo* at 800° C. for 0 5 hr.								
15	0·299	Very slow	1·52	1·425	1·46	0 0190	0 0145	0 0935
13	0·300	R.R.	1·625	1·58	1 65	0 0214	0 0168	0·1218
16	0·299	R.R.	1·68	1 60	1 61	0 0186	0·0129	0·1190
14	0·299	4	1·73	1·655	1·68	0·0209	0·0149	0·1440
Heated *in vacuo* at 900° C. for 0·5 hr.								
12	0·298	Very slow	1 299	1·285	1·31	0·0157	0·0015	0·0541

TABLE V. STEEL NO. 11 O. PULLED IN STAGES WITH RESTS. GEAR R.R. ORIGINAL DIAMETER, 0·399 IN. DIAMETERS NOT MEASURED AT RESTS

	Load (tons)						Extension (in.)
Ext. no.	Top of peak	Bottom	End of ext.	After rest	Increase after rest	Time of rest hr.	Total from zero
1	1·69	1·47	1·90	1·81	—	—	0·2446
2	2·085	1·98	2·16	2·05	0·27	0·5	0·362
3	2 32	2 21	2·31	2·21	0·27	0·5	0·473
4	2·47	2·37	2·44	2 32	0 26	0·5	0 598
5	2 58	2·48	2·52	2·35	0·26	0·5	0·711
6	2·845	2·62	2·62	2·51	0·49	2·5	0·820
7	2·79	2·66	2·65	2 585	0·28	1	0·935
8	3·08	2·90	Curve falling		0·50	17	

TABLE V. STEEL NO. 13 T. HEATED *IN VACUO* AT 800° C. PULLED IN STAGES WITH RESTS. GEAR R.R. ORIGINAL DIAMETER, 0·300 IN.

	Load (tons)						
Ext. no	Top of peak	Bottom	End of ext.	After rest	Increase after rest	Time of rest hr.	Extension (in.) Total from zero
1	1·625	1·58	1·865	1·75	—	—	0·1830
2	2·04	1·82	2·10	1·98	0·29	1	0·2870
3	2·14	No drop	2·21	2·06	0·16	1	0·3664
4	2·295	2·02	2·33	2·22	0·23	23	0·5039

TABLE VI ARMCO IRON NO. 11_2 HEATED *IN VACUO* AT 950° C. FOR 0·5 HR PULLED IN STAGES WITH RESTS. GEAR R.R

		Load (tons)							
Ext. no	Area sq. in.	Top of peak	Bottom	End ext.	After rest	Increase after rest	Increase in stress tons/sq in	Time of rest hr.	Extension (in.) Total from zero
1	0·122	1·20	1·10	1·46	1·385	—	—	—	0·1486
2	0·117	1·765	1·595	1·83	1·75	0·38	3·11	0·5	0·2932
3	0·112	2·165	2·065	2·18	2·09	0·41	3·50	0·5	0·5155
4	0·109	2·50	2·31	1·91	2·28	0·41	3·66	0·5	0·7480
5	0·105	2·98	2·79	2·84	2·75	0·70	6·66	22·5	0·9601

TABLE VI. STEEL NO. 12 O. PULLED IN STAGES WITH RESTS. GEAR NO. 3

		Load (tons)							
Ext. no	Area sq. in.	Top of peak	Bottom	End of ext.	After rest	Increase after rest	Increase in stress tons/sq. in.	Time of rest hr.	Extension (in.) Total from zero
1	0·1255	1·52	1·48	1·985	1·875		—	—	0·2758
2	0·1208	2·27	2·06	2·38	2·255	0·40	3·31	1	0·5027
3	0·116	2·55	2·44	2·58	2·44	0·30	2·58	0·5	0·7354
4	0·111	2·71	2·61	2·67	2·53	0·27	2·43	0·5	0·9725
5	0·1075	2·80	2·70	2·72	2·585	0·27	2·51	0·5	0·1061

TABLE VI. ARMCO IRON NO. 12_2. HEATED *IN VACUO* AT 950° C. PULLED IN STAGES WITH RESTS. GEAR R.R.

		Load (tons)							Extension (in.)
Ext. no	Area sq. in.	Top of peak	Bottom	End of ext	After rest	Increase after rest	Increase in stress tons/sq. in.	Time of rest hr.	Total from zero
1	0·1245	0·205	1·13	1·46	1·395	—	—	—	0·1355
2	0·122	1·77	1·595	1·75	1·655	0·38	3·11	0·5	0·2184
3	0·120	2·73	2·12	2·22	2·13	1·08	9·00	17·5	0·349
4	0·1175	2·54	1·90	2·445	2·345	0·41	3·49	0·5	0·515
5	0·115	2·72	2·56	2·56	2·47	0·38	3·30	0·5	0·689
6	0·113	2·80	2·61	2·655	2·56	0·33	2·92	0·5	0·833
7	0·1105	2·87	2·72	2·74	2·60	0·31	2·81	0·5	0·966
8	0·106	3·17	3·00	2·91	2·75	0·57	4·90	22	—

(Load falling)

PART III. TENSILE TESTS ON ALUMINIUM ALLOYS

In view of the fact that duralumin was reputed to have a yield point similar to iron and steel and to show Luder's Lines on straining (Dawidenkow 1930), some tests were made on both duralumin and some High Duty alloys of different heat treatment.

A quenched and aged duralumin gave a uniform, smooth curve. One that was tested immediately after quenching gave a curve similar to that illustrated in fig. 13. At a certain point in the extension, it seemed to become unstable and slipped in large jumps which increased as the test progressed. At the same time the bar became uneven. If the test-piece were kept in liquid air after quenching until it could be tested, the jerky deformation began sooner. It was thought possible that the process of age hardening might be responsible for the jumping deformation so tests were made on certain High Duty aluminium alloys of different heat treatment. Some of these are stable at ordinary temperatures even in the quenched state, but the quenched alloys did not show this type of deformation. Both of those which developed jumps in the course of the test were annealed and slowly cooled (Nos. 1, 5, Table VII, figs. 12, 13).

Some of the alloys were also tested at two rates of straining. Particulars of results are given in Table VII.

The yield points were indefinite except in No. 5 so that these figures are only approximate. In every case, the curve at the slow speed slightly exceeded the fast, in this way resembling armco iron. It is known (Guillet 1926, Teed 1936) that deformation assists age hardening by precipitation in quenched duralumin but there is no reason to expect precipitation in annealed alloys. On the other hand, tests on aluminium at two speeds at different temperatures carried out by Martin (1924) always gave a lower result for the slow speed, in this way resembling copper.

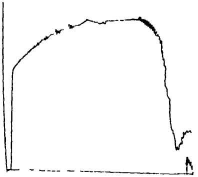


FIG. 12. No. 5_1 gear 3. High Duty alloy R.R. 72.

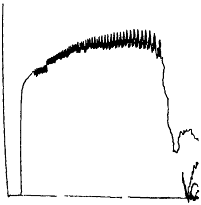


FIG. 13. No. 5_2, very slow. High Duty alloy R.R. 72

TABLE VII. TENSILE TESTS ON HIGH DUTY ALLOYS

Original Diam. 0·35 in. Length 5 in.
Fast = gear No. 3 } same as in previous tests, Table II.
Slow = very slow speed }

	No. 1		No. 2		No. 3		No. 5	
	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow
Stress, tons/sq. in. at yield	5 32	6·76	9·09	10·00	7·78	7·61	11·25	11·75
Maximum stress	17·10	17·70	22·70	23·4	23·40	24·20	18·15	19·60
Extension* % on 5 in.	8·86	10 25	16·10	15·85	17·80	17 65	9·65	10·15

* The extension does not include fracture.

R.R. 56
1. annealed, furnace cooled, soaked, 330° C.
2. annealed, air cooled, soaked, 330° C.
3. soaked, 525° C., quenched.

R.R. 72
5. annealed.

It is characteristic of the deformation by jumps that their magnitude increases with the stress. The process is, in fact, similar to the increase of the drop in load illustrated in fig. 8 in the case of the copper test-piece which was stopped for half a minute when the rate of extension was changed. The faster gear, which was responsible for the increased stress, gave a greater drop. The release of the spring caused by the sudden extension associated with each jump in the aluminium alloys causes the pointer to swing back and to vibrate and mask the curve, but the peaks represent the loads held by the test-piece just before each jump occurs and the distance between them the amount of strain associated with each.

The lower series of points is unreliable for the purposes of measurement owing to the inertia of both spring and pointer.

In the duralumin, the maximum extension of each jump near fracture amounted to 0·0165 mm.; in the High Duty alloy No. 5 to 0·0384 mm. Each step in the extension left the metal harder than it was before the load immediately began to rise again and a line drawn through the tops of the peaks also gave a gradually rising curve.

The influence of rate of extension on the development of the jumps can be compared in figs. 12, 13. They reach their maximum at the slow speed. In this respect they may be compared with the rise and fall in the curves following a rest in iron and steel. In fact, the only difference appears to be that the jumps are not repeated in iron and steel except at very slow rates of straining (see figs. 5, 6).

The observations recorded above agree with the results of previous investigators on the occurrence of this type of deformation and it may well be that all plastic distortion takes place by steps and is not a continuous process. The uniformity of both dimensions and frequency of occurrence which has been noted in zinc and cadmium (Becker and Orowan 1932; Schmid and Valouch 1932) and also rock-salt crystals (Classen-Nekludowa 1929, Dawidenkow 1930; Joffé 1928) suggests that they are related to the crystal structure On the other hand, it has also been shown that they follow damage to a crystal (Becker and Orowan 1932, Orowan 1934); for example, if a zinc crystal is bent or, if in course of preparation, it is drawn out of the molten bath too quickly.

The evidence is in favour of regarding a jump as a result of slip interference, by which the stress increases abnormally. When slip finally recommences the rate of deformation is relatively so great that a large slip occurs before the metal hardens. The effect of rate of strain on the magnitude of the jumps is twofold (1) the hardening at slow speeds is greater in these alloys than at fast speeds, therefore the peaks tend to be higher; (2) the slow speed ensures that the full extension of the test-piece is recorded in the movement of the spring.

PART IV. DISCUSSION OF RESULTS

Previous work has been confirmed in all the experiments undertaken. With a uniform method of testing, the rate of deformation stands out as one of the most important factors influencing the stress-strain curve in both ferrous and non-ferrous metals. It determines the stress at which plastic yielding commences. This in turn, together with the rate of extension and flexibility of the spring, determines the value of the lower yield in iron and steel. At sufficiently slow rates of extension there may be only one yield point and no horizontal part of the stress-strain diagram. A high yield stress gives a longer yield. Hence the influence of grain size on length of yield is also explained, as a fine-grained metal has a higher yield stress than a coarse. The mechanism of the drop at the yield appears to be similar to that associated with the jumping deformation described in Part III. The difference lies in the fact that in iron and steel the load does not rise again to the same value as the primary yield until a considerable further extension has taken place, whereas in the aluminium alloys the rise follows immediately. This particular type of slip is associated with a rise in stress above normal due to two principal causes· (1) high speed of loading, which suggests that the load must be applied for a certain minimum time before the metal begins to flow, (2) an internal slip hindrance which seems to be connected with distortion of the crystal structure.

The extension at the yield point in iron and steel has been shown to be very localized, so that large deformations take place in a relatively small number of crystals instead of being uniformly distributed over the whole length of the test-piece. This must cause instability in stress distribution and a tendency towards further deformation at a lower load, which is sometimes observed through the yield, although there is more often a slight rise. It is generally accepted that any distortion of the crystal structure increases the resistance to further deformation and the material within a Lüder's Line has been shown to be harder than the unstrained metal. Therefore, immediately one part is deformed it will cease to yield further while the rest is undistorted, after which the stress must be increased to deform it further. If the steps by which this change proceeds are sufficiently small and are uniformly distributed, a gradually rising stress-strain curve is the result. In the case of the initial yield of iron and steel, the deformation, having once begun, proceeds normally during the horizontal part of the curve. But it is more difficult to explain how a large local extension can take place in the first place without affecting neighbouring crystals by causing an immediate hardening and spread of the distortion.

A comparison between the fast and slow rates of extension and the effect of rests on the different metals suggest a possible explanation. For the same extension, a very slow rate of deformation produces a greater degree of hardening in iron and low carbon steels and certain aluminium alloys and a lesser degree of hardening in copper, a higher carbon steel and aluminium. At normal and higher rates of deformation there is an increase of stress with velocity of deformation for all these metals, which has already been investigated by Ludwik (1909) and Deutler (1932), and the latter has succeeded in confirming Prandtl's (1928) theory relating thereto. Prandtl assumes that certain atoms of a slip-plane in course of gliding are in an unstable position and require time to reach new positions of equilibrium. This may account for elastic after-working and a diminution of stress when deformation ceases (see p. 580). If sufficient time elapses, the metal may undergo a further process of recovery and may even soften considerably and creep under load. The process is accelerated by heat and if the temperature is high enough may counterbalance the effects of work-hardening. On the other hand it is reasonable to assume that if equilibrium is not reached immediately during deformation, the after effects may result equally well in an increase in hardness. On this hypothesis, a metal deformed at a slow rate may harden more for a given deformation than if it is deformed rapidly and a rest following a rapid distortion will enable the metal to reach its equilibrium state. This is a possible explanation of the experimental results obtained and is further confirmed by the fact that the increase in hardness is facilitated in iron and steel by low temperature annealing and that the first effects of heat on cold worked metals is often to harden them (Mathewson and Phillips 1916).

As increase in hardness implies an increase of potential energy, it is to be expected that other physical properties will show a change in the same direction. There is no evidence to confirm this at present.

Assuming that certain metals do not immediately reach their maximum hardness on straining, a high rate of deformation may result in a proportionately large distortion before the resistance to further distortion becomes effective. While velocity of deformation determines the stress, the stress in turn determines the rate of deformation under the influence of the spring in Polanyi's and Quinney's machines.

The views expressed above are in close agreement with those of Orowan* (Orowan 1934, 1935*a*, *b*, 1936) on the influence of time in testing. He also, arrives at the conclusion that rest can equally well cause recovery or an increase in resistance to deformation. There may be a critical rate of straining for each metal at a given temperature at which rate of increase of resistance to shear is balanced by rate of recovery.

---

* A critical discussion of some of Orowan's work is to be found in the "Report on Viscosity and Plasticity" by W. G. and J. M. Burgers (1936).

If the rate of deformation is so high that the metal does not harden, the flow is proportional to the stress and, in the case of the horizontal part of the curve of iron and steel, it closely approaches true plastic distortion.

One point remains to be discussed. These experiments do not throw any more light on the problem of the high initial stress required to cause the first plastic yield in iron and steel or the rise which follows a rest. There seems to be greater difficulty in starting than in continuing plastic deformation. Orowan refers to "slip-hindrance" in general terms and connects it with dislocations in the crystal lattice. There is, however, a close relation between rate of loading and stress and it appears that a load must be applied for a certain minimum time before flow commences.

It is clear from the foregoing experiments that the rate of application of load should be stated in all tests, particularly when an autographic recording machine of the spring type is used. In fact, it may be said that no test is complete unless stress, strain, time and temperature are all recorded.

This work was carried out in the Engineering Department, Cambridge University. I wish to express my thanks to Professor C. E. Inglis, F.R.S., for the facilities which have been given me in his laboratory. I am particularly indebted to Mr G. S. Gough, M.A., for suggestions and advice which have proved invaluable.

REFERENCES

Andrade, E. N. da C. 1910 *Proc. Roy. Soc.* A, **84**, 1.
— 1914 *Proc. Roy. Soc* A, **90**, 329.
Bach, C. 1904 *Z. Ver. dtsch. Ing.* **48**, 1040.
— 1905 *Z. Ver. dtsch. Ing.* **49**, 615.
Becker, R. and Orowan, E. 1932 *Z. Phys.* **79**, 566.
Bengough, G. D. and Hudson, O. F. 1910 *J. Inst. Met.* **4**, 92.
Bernhardt, E. O. 1936 *Metallwirtschaft*, **15**, 889.
Boas, W. and Schmid, E. 1930 *Z. Phys.* **61**, 767.
Braunbek, W. 1929 *Z. Phys.* **57**, 501.
Burgers, W. G. and J. M. 1936 "First Report on Viscosity and Plasticity." *Verh. Akad. Wet., Amst.*, Afd. Natuurk., **15**, No. 3.
Classen-Nekludowa, M. 1929 *Z. Phys.* **55**, 555.
Cook, G. 1931 *Philos. Trans.* A, **230**, 103.
Dawidenkow, N. 1930 *Z. Phys.* **61**, 46.
Deutler, H. 1932 *Phys. Z.* **33**, 247.

Docherty, I. G. and Thorne, F. W. 1931 *Engineering*, **132**, 295.
Edwards, C. A. and Pfeil, L. B. 1925 *J. Iron Steel Inst.* **112**, 79.
Elam, C. F. 1927 *Proc. Roy. Soc.* A, **116**, 694.
Emsslin 1928 *Z. Ver. dtsch. Ing.* **72**, 1625.
Foll, E. W. 1927 *Carnegie Schol. Mem* **16**, 101.
— 1935 *J. Iron Steel Inst.*
Ginns, D W. 1937 *J. Inst. Met.* **59**, 61.
Guillet, L. 1926 *Rev. Met.* **23**, 48.
Hopkinson, B. 1905 *Proc. Roy. Soc.* A, **74**, 498.
Joffé, A. 1928 "Physics of Crystals," p. 50. New York.
Korber, F. and Pomp, A. 1934 *Mitt. K.-Wilh -Inst. Eisenforsch.* **16**, 179.
Kuhnel 1928 *Z. Ver. dtsch. Ing.* **72**, 1226.
Kuntze, W. and Sachs, G 1928 *Z Ver. dtsch. Ing.* **72**, 1011.
Ludwik, P. 1909 *Phys. Z.* **10**, 411.
Martin, T. 1924 *J. Inst. Met.* **31**, 121.
Mathewson, C. H. and Phillips, A 1916 *Bull. Amer. Inst. Min. (Metall.) Engrs*, No 1.
Moser, H. 1928 *Stahl u. Eisen, Dusseldorf*, p. 1601.
Muir, J 1900*a* *Philos. Trans* A, **193**.
— 1900*b* *Philos. Trans.* A, **198**.
Nadai, A. 1931 "Plasticity." New York.
Orowan, E. 1934 *Z. Phys* **89**, 605.
— 1935*a* *Z. Phys.* **97**, 573.
— 1935*b* *Z. Phys* **98**, 382
— 1936 *Z. Phys* **102**, 112
Polanyi, M. 1925 *Z. tech Phys.* **6**, 121.
Pomp, A. and Krisch, A. 1937 *Mitt K -Wilh -Inst Eisenforsch.* **19**, 330.
Prandtl, L. 1928 *Z. angew. Math. Mech* **8**, 85.
Quinney, H. 1934 *Engineer, Lond* , **157**, 332
— 1936 *Engineer, Lond.*, **161**, 669
Robertson, A. and Cook, G 1913 *Proc. Roy. Soc* A, **88**, 462.
Sachs, G and Shoji, H 1921 *Z Phys.* **45**, 776.
Schmid, E. and Boas, W 1935 "Kristallplastizitat", p 125. Berlin.
Schmid, E. and Valouch, M A. 1932 *Z Phys.* **75**, 531.
Siebel, E and Pomp, A. 1928 *Mitt. K -Wilh -Inst Eisenforsch.* **10**, 63.
Siebel, E and Schwaigerer, S 1937 *Metallwirtschaft*, **16**, 701
Späth, W. 1937*a* *Metallwirtschaft*, **16**, 193.
— 1937*b* *Metallwirtschaft*, **16**, 697.
Teed, P. L. 1936 *J Inst. Met.* **58**, 48
Welter, G. 1935 *Metallwirtschaft*, **14**, 1043.
— 1936 *Metallwirtschaft*, **16**, 885.

INDEX TO VOLUME CLXV (A)

Adsorption of vapours at plane surfaces (Bangham and Mosallam), 552.
Adsorption phenomena, oscillography (Johnson and Henson), 148.
Archer (C. T.) Thermal conduction in hydrogen-deuterium mixtures, 474
Arnot (F. L.) and M'Ewen (Marjorie B.) The formation of mercury molecules, 133.
Atkins (W. R. G.) Photo-electric measurements of the seasonal variations in daylight around $0\cdot41\mu$, from 1930 to 1937, 453.

Bangham (D. H.) and Mosallam (S.) The adsorption of vapours at plane surfaces of mica Part 1, 552
Benger (M.) *See* Randall, Benger and Groocock
Blackett (P. M. S.) The nature of the penetrating component of cosmic rays, 11
Blackett (P. M. S.) and Wilson (J. G.) The scattering of cosmic ray particles in metal plates, 209.
Born (M.) A suggestion for unifying quantum theory and relativity, 291.
Bretscher (E.) *See* Feather and Bretscher.
Burhop (E. H. S.), Hill (R. D.) and Townsend (A. A.) The production of gamma-rays by neutrons, 116.

Coal, chemical constitution (Randall, Benger and Groocock), 432.
Corin (C.) and Sutherland (G. B. B. M.) The infra-red absorption spectrum of methylene chloride, 43.
Cosmic rays (Blackett), 11.
Cosmic ray particles, scattering (Blackett and Wilson), 209.
Cosmology (Dirac), 199.
Coulson (C. A.) and Duncanson (W. E.) Comparison of wave-functions for HeH^{++} and HeH^{+}, 90
Crystal beams of sodium-ammonium seignette salt (Mandell), 414.

Devonshire (A. F.) *See* Lennard-Jones and Devonshire.
Dirac (P. A. M.) A new basis for cosmology, 199.
Duncanson (W. E.) *See* Coulson and Duncanson.

Elam (C. F.) The influence of rate of deformation on the tensile test with special reference to the yield point in iron and steel, 568.
Electromagnetism equations (Milne), 313, 333.
Electron ferromagnetism (Stoner), 372.

Fage (A.) The influence of wall oscillations, wall rotation, and entry eddies, on the breakdown of laminar flow in an annular pipe, 501.
Fay (J. W. J.), Glückauf (E.) and Paneth (F. A.) On the occurrence of helium in beryls, 238.
Feather (N.) and Bretscher (E.) Uranium Z and the problem of nuclear isomerism, 530.
Ferromagnetism, collective electron (Stoner), 372.
Fluid motion (Schmidt and Saunders), 216.

Foster (J. S.), Langstroth (G. O.) and McRae (D. R.) Quantitative spectrographic analysis of biological material. III. A method for the determination of sodium and potassium in glandular secretions, 465

Gamma-rays, production by neutrons (Burhop, Hill and Townsend), 116.
Gases, critical phenomena (Lennard-Jones and Devonshire), 1.
Glandular secretions, sodium and potassium in (Foster, Langstroth and McRae), 465.
Glückauf (E.) *See* Fay, Glückauf and Paneth.
Glückauf (E.) and Paneth (F A) Identification and measurement of helium formed in beryllium by γ-rays, 229.
Goodeve (C. F) *See* Porret and Goodeve.
Gough (H. J.) and Wood (W. A.) The crystalline structure of steel at fracture, 358.
Groocock (C. M.) *See* Randall, Benger and Groocock.

Helium in beryls (Glückauf and Paneth; Fay, Gluckauf and Paneth), 229, 238.
Henson (A. F.) *See* Johnson and Henson
Hill (R. D.) *See* Burhop, Hill and Townsend.
Hyperfine structure (Jackson and Kuhn), 303.

Jackson (D. A.) and Kuhn (H) Hyperfine structure, Zeeman effect and isotope shift in the resonance lines of potassium, 303.
Jeffreys (H.) Significance tests when several degrees of freedom arise simultaneously, 161.
Johnson (M. C) and Henson (A F) Oscillography of adsorption phenomena. III. Rates of deposition of oxygen upon tungsten, 148.

Kothari (D S.) The theory of pressure-ionization and its application, 486.
Kuhn (H.) *See* Jackson and Kuhn.

Langstroth (G. O.) *See* Foster, Langstroth and McRae.
Lennard-Jones (J. E.) and Devonshire (A. F.) Critical phenomena in gases. II. Vapour pressures and boiling points, 1

Mandell (W.) Resonance in crystal beams of sodium-ammonium seignette salt, 414.
McRae (D. R.) *See* Foster, Langstroth and McRae.
Mercury molecules, formation (Arnot and M'Ewen), 133.
M'Ewen (Marjorie B.) *See* Arnot and M'Ewen
Milne (E. A.) On the equations of electromagnetism. I. Identifications, 313
Milne (E. A.) On the equations of electromagnetism. II. Field theory, 333.
Mosallam (S) *See* Bangham and Mosallam.

Neutrino theory of light (Pryce), 247.

Paneth (F. A.) *See* Glückauf and Paneth, *and* Fay, Glückauf and Paneth.
Photo-electric measurements of daylight (Atkins), 453.
Porret (D.) and Goodeve (C. F) The continuous absorption spectra of alkyl bromides and their quantal interpretation, 31.

Pressure-ionization (Kothari), 486.
Price (W. C.) and Simpson (D. M.) The absorption spectra of sulphur dioxide and carbon disulphide in the vacuum ultra-violet, 272.
Pryce (M. H. L.) On the neutrino theory of light, 247.

Quantum theory and relativity (Born), 291.

Randall (R. B.), Benger (M.) and Groocock (C. M.) The alkaline permanganate oxidation of organic substances selected for their bearing upon the chemical constitution of coal, 432

Salter (C.) *See* Simmons and Salter
Saunders (O. A.) *See* Schmidt and Saunders.
Schmidt (R. J.) and Saunders (O. A.) On the motion of a fluid heated from below, 216.
Significance tests (Jeffreys), 161.
Simmons (L. F. G.) and Salter (C.) An experimental determination of the spectrum of turbulence, 73.
Simpson (D. M.) *See* Price and Simpson.
Spectra, absorption, of alkyl bromides (Porret and Goodeve), 31.
Spectra, absorption (Price and Simpson), 272.
Spectrographic analysis of biological material (Foster, Langstroth and McRae), 465.
Spectrum of methylene chloride (Corin and Sutherland), 43.
Spectrum of turbulence (Simmons and Salter), 73.
Steel, crystalline structure (Gough and Wood), 358.
Stoner (E. C.) Collective electron ferromagnetism, 372.
Sutherland (G. B B M) *See* Corin and Sutherland.

Tensile test with special reference to the yield point (Elam), 568.
Thermal conduction in hydrogen-deuterium mixtures (Archer), 474.
Tomotika (S.) Application of the modified vorticity transport theory to the turbulent spreading of a jet of air, 65.
Tomotika (S.) On the velocity and temperature distributions in the turbulent wake behind a heated body of revolution, 53
Townsend (A. A) *See* Burhop, Hill and Townsend
Turbulent spreading of a jet of air (Tomotika), 65.

Uranium Z (Feather and Bretscher), 530

Velocity and temperature distributions (Tomotika), 53.

Wave-functions for HeH^{++} and HeH^{+} (Coulson and Duncanson), 90.
Wilson (J. G) *See* Blackett and Wilson.
Wood (W. A.) *See* Gough and Wood.

END OF THE ONE HUNDRED AND SIXTY-FIFTH VOLUME (SERIES A)

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

3 MARCH 1938

Investigations of the mechanism of the transmission of plant viruses by insect vectors. II. By H. H. STOREY. (*Communicated by F. T. Brooks, F.R.S.—Received 26 January* 1938.)

An attempt has been made to determine the conditions that decide whether *Cicadulina mbila* succeeds or fails in transmitting the virus of streak disease of maize. Evidence shows that individual insects, although belonging to an active race and therefore all capable of acting as vectors, vary greatly in their ability to cause infection during a short period of contact with the plant. The results obtained by comparing the infections following contacts of single insects and groups of insects are interpreted as conforming with an hypothesis of independence; that the effect of one insect is independent of the effects of any other insects that may puncture the plant. If this interpretation be correct, a small but definite variation in susceptibility of the maize plants used can be recognised.

By removing at the end of the contact the portion of leaf-tissue exposed to the insect, the probability of infection was somewhat reduced. The effect, though probably significant, was small; and it was concluded that the virus had normally become established in the plant and had moved down at least a few millimetres during the period that the insect maintained contact.

Cicadulina mbila, while resting on a maize leaf, always has its mouthparts inserted, although it may change the position of puncture and may suck material from the leaf only intermittently The stylets penetrate all tissues of the leaf, but appear to be moved frequently until the phloem is entered.

The insect can take up virus from a chlorotic area of a diseased leaf during a puncture lasting only 15 seconds, which never penetrates beyond the mesophyll. If confined to the green part of the leaf lying between the chlorotic areas, punctures whether to the mesophyll or to the phloem, fail to take up any virus

No feature in the behaviour of the insects while in contact with a leaf could be certainly related to their success in causing infection, except that the observations suggested that a high defaecation rate is unfavourable. Evidence was obtained that a single puncture can result in infection The traces of all such single infective punctures examined entered the phloem; those of unsuccessful punctures sometimes entered the phloem and sometimes ended elsewhere.

The insect can inoculate the virus successfully by inserting its stylets through a wax membrane into a leaf held below, but only if the membrane is not too thick to prevent the stylets reaching the phloem. Seedlings were not infected by insects feeding on the coleoptile; in this organ the phloem is deep-seated, beyond the reach of the stylets.

Plants were never infected by punctures, however numerous, if they were all of less than a certain duration This threshold-period is about 5 min. at temperatures between 23 and 26° C. A study of the puncture traces showed that many sub-threshold punctures penetrated to the phloem

A consideration of all the evidence causes me to advance the hypothesis that the insect inoculates the virus in distinct doses, each independent in its effect of any other doses that may be inoculated by the same or other insects. The delivery of a dose is determined by some incident that occurs only after puncture has been maintained for some time.

The influence of rate of deformation on the tensile test with special reference to the yield point in iron and steel.

By C. F. Elam. *(Communicated by Sir Harold Carpenter, F.R.S. Received 26 January 1938.)*

Tensile tests on armco iron and two steels have been carried out in a 50 ton Buckton testing machine and in a spring testing machine designed by Quinney The distortion at the yield point and the formation of Luder's Lines were investigated in both machines Tests with these metals and with copper and certain aluminium alloys at different speeds confirmed previous observations that the faster the rate of loading the higher the stress. On the other hand, at very low speeds, there was a greater increase of hardness in the case of iron and mild steel and the aluminium alloys than at faster rates of testing In copper the reverse was the case The deformation at the yield point is compared with the "jumping" deformation characteristic of many substances. It is suggested that under certain conditions, the rate of deformation may be greater than the rate of increase in hardness and that rests or very slow rates of extension may equally well increase the effect of work-hardening as allow relaxation to take place

The resistance of superconducting cylinders in a transverse magnetic field.

By A. D. Misener *(Communicated by J. D. Cockroft, F.R.S.—Received 31 January 1938.)*

Detailed experiments on the transition of polycrystalline cylindrical wires from the superconducting to the normal state in a uniform transverse magnetic field are described Several specimens each of tin, indium and lead of high purity were investigated. The applied field strength (H_t) which restored the first traces of electrical resistance was accurately determined at various temperatures below the normal transition point. It is found that the ratio of this field to the critical field (H_k) corresponding to the temperature of the experiment is not constant but varies with the temperature. The variation is linear, the ratio H_t/H_k increasing as the temperature is decreased. For all specimens the ratio H_t/H_k would have the value 0 50 at the normal transition point.

It is considered that this effect is incompatible with the idea of an "intermediate state" existing in a long cylinder, and an explanation is offered based on the assumption that after penetration of the external magnetic field (above 0 50 H_k) the cylinder breaks up into macroscopic regions of normal and superconducting material. The agreement of this hypothesis with previously observed phenomena is discussed

A preliminary experiment is reported in which, by measuring the resistance of successive millimeter sections of a specimen, the coexistence of these normal and superconducting regions is shown directly

---

β-transitions in a coulomb field. By F. Hoyle. (*Communicated by R. H. Fowler, F.R.S.—Received* 31 *January* 1938.)

This paper attempts to give the selection rules, and the possible forms of the electron energy spectra, which correspond to elements on the first, second, and third Sargent curves, in the case of each of the possible forms of β-interaction belonging to Hamiltonians that contain a derivative of only the neutrino wave function. This allows a choice of several possibilities for the interaction, among which is the form proposed by Konopinski and Uhlenbeck. The accurate solution of the problem would require a knowledge of the wave equation of a nucleus containing many particles I assume that a non-relativistic Schrodinger equation can be formulated for the nucleus, in which the spin co-ordinate of each particle has two possible values (3, 4). The solutions of such equations for the initial and final nuclei will give a first type of forbidden transition It is further assumed that a relativistic equation can be constructed for the nuclei, in which the spin co-ordinate of each particle will now have four values (1, 2, 3, 4). From this point of view I regard the Schrödinger equations as given by neglecting, in the relativistic equations, all components of the wave function in which any spin co-ordinate is 1, 2. Generalizing from the Pauli reduction of the Dirac equation in the single-body problem, to the assumption that a component of the wave function in which one spin co-ordinate is different from 3, 4 can be expressed to a suitable approximation, in terms of those components which occur in the Schrodinger equation, the connexion between these components seems to be analogous to that given by Pauli in the one-body problem. This introduces small components of the wave functions into the expression for the transition probability. These small components will have selection rules which are different from those for the large components, and transitions which were forbidden may now become "allowed" for the small components (that is, the light particle wave functions may be treated as constants over the nucleus). It is convenient to distinguish those components which are small in the spin variable of the transition particle, and those which are small in other variables. These two groups of small components will, in general, also have different selection rules. The result of comparing the selection rules for the three groups of components of the wave functions (the large components, and the two groups of small components), and of discussing the corresponding forms of the electron spectra, show that there may be elements on the second Sargent curve with either

(1) The "allowed distribution form" given by Fermi,
or possibly for light elements (nuclear charge $\leqslant 20$).

(2) Electron distributions which differ from (1), and of types previously discussed (Hoyle 1937; p. 290, fig. 4, I, II),
and that for elements on the third Sargent curve we have the possibilities

(1) The "allowed distribution form" given by Fermi.

(2) Effectively the distribution (1), but with a slight humping at the upper and lower energy limits.

(3) Distributions of the types previously discussed (Hoyle 1937; fig. 4, I, II).

(4) And for light elements (nuclear charge $\leqslant 20$) distributions which differ more widely from the "allowed form" than (3), the shapes of these distributions being similar to (3), but of a more exaggerated form.

On the nuclear forces and the magnetic moments of the neutron and the proton.

By H. Frohlich, W. Heitler and N. Kemmer. *(Communicated by N. F. Mott, F.R.S.—Received 1 February 1938.)*

An attempt is made to explain the properties of the nuclear particles proton and neutron by the hypothesis that these particles are capable of emitting a positive or negative "heavy electron" respectively with a rest mass m_0 between that of the proton and the electron. The existence of those particles has been made probable by cosmic ray observations.

The heavy electrons are assumed to have no (or integral) spin and satisfy Bose-statistics. The wave functions of these particles are assumed to be of vectorial character, the components of which satisfy the Klein-Gordon equation. They are quantised according to the scheme given by Pauli and Weisskopf. Thus, a free heavy electron can exist in three different states of polarization, there are two transverse and one longitudinal wave with given momentum. The interaction with the nuclear particles is found by relativistic arguments and contains—apart from the mass m_0—two arbitrary constants g and f, both with dimensions of an electric charge.

With this scheme we have calculated:

1. The neutron-proton force It is an exchange force and has a range $1/\lambda = \hbar/m_0 c$. In the ^{3}S-state the force is always attractive. (This would not be the case for a scalar wave function) g and f can be chosen so that the ^{3}S- and ^{1}S-states have the right position. The experiments suggest $g \doteqdot f \doteqdot 5$ electron charges.

2. The magnetic moments μ_P and μ_N of the proton and neutron. They are found to be of nuclear dimensions and have the right sign.

3. The mass m_0 can be determined independently from the range of the neutron-proton force and from the magnetic moments. In both ways we find $m_0 \doteqdot 100$ electron masses.

4. The proton-proton force is obtained only in the fourth order of approximation and leads to a strong repulsion for distances less than $1/2\lambda$. Attraction and equality with the neutron-proton force could be attained by introducing also neutral particles with mass m_0.

5. The theory leads to a diverging self energy of the proton and neutron

---

Photoelectric measurements of the seasonal variations in daylight around 0·41μ from 1930 to 1937.

By W. R. G. Atkins, F.R.S. (*Received* 1 *February* 1938.)

In collaboration with H. H. Poole photoelectric cells were standardized and records obtained with a vacuum sodium cell and Cambridge "thread recorder" for 1930 (*Philos. Trans.* 1935, **235**, 1–27; 1936, **235**, 245–72). Further measurements are now summarized and compared with those for 1930, after establishing the constancy of the cell used. The greatest, least and mean values of the daily maximum, in kilolux, are given for each month, also corresponding values for the total vertical illumination in kilolux hours. For the sake of uniformity all the daily curves were measured by the author.

Five years were very similar and averaged 309 kilolux hours a day, but 1934 gave 350 and 1930 gave 414. The greatest sunshine average was 5 17 hr. in 1933. A comparison with the meteorological returns failed to explain satisfactorily the high values for 1930, though radiation at South Kensington was rather high that year. It is possible that specially clear upper air was prevalent in 1930, or it may be that the explanation lies in a variation in the ratio of ultra-violet to green in the solar spectrum, as suggested by Pettit but since contested by Bernheimer.

The coagulation of plasma by trypsin.

By J. Mellanby, F R.S and C. L. G. Pratt. (*Received* 2 *February* 1938.)

Trypsin digests the fibrinogen contained in oxalated plasma when the amount added is greater than the quantity of antitrypsin in the plasma. Stable, non-oxalated, fowl plasma is subject to the same effect, but certain concentrations of trypsin bring about coagulation. This coagulant action is due to the liberation by trypsin of thrombokinase previously masked in the plasma. Contrary to certain recent claims, trypsin does not activate prothrombase. It is possible to prepare fowl plasma, free from thrombokinase, such that the addition of thrombokinase brings about coagulation whereas the addition of trypsin does not The results indicate that thrombase is not identical with trypsin and that thrombokinase and calcium cannot be replaced by trypsin in the activation of prothrombase.

Studies on the hypophysectomized ferret. X. Growth and skeletal development.

By A. S. Parkes, F.R.S. and I. W. Rowlands. (*Received* 4 *February* 1938.)

Six immature male and two immature female ferrets were hypophysectomized at a body weight of 500–600 g. (7–10 weeks of age). Subsequent body growth in ferrets of both sexes was retarded; in the female the stasis was almost complete but the males continued to grow until the age when the growth in the normal ferret is complete.

Skeletal development was studied by means of X-ray photographs and prepared skeletons. It was found that skeletal growth had been arrested; the bones had a low calcium content and open epiphyses. Stasis in the development of the skull was most marked, this being shown by the absence of sagittal and nuchal crests, open sutures and the retention of all the characteristics of the skull of an immature animal.

The reaction between oxygen and nitric oxide.

By E. M. Stoddart. (*Communicated by F. G. Donnan, F R S.—Received* 4 *February* 1938.)

It has been shown that the observations of Baker on the non-interaction between intensively dried oxygen and nitric oxide can be repeated, provided that care is taken to ensure that mixing takes place entirely in the oxygen-containing bulb. The only explanation of this phenomenon is that the reaction between the gases is a heterogeneous one, an explanation which fits the observations of Hasche, who showed that the rate of this reaction was lowered by as much as 20% in paraffin coated vessels. The present author suggests that the drying has no effect except perhaps that of removing the adsorbed water film from the glass vessels, thus allowing the surfaces to adsorb a complete gas film in its place. When this gas film is oxygen, no reaction occurs in the mixed gases, but when the gas film is nitric oxide, reaction is possible. At least this explanation agrees with the known facts regarding heterogeneous reactions.

With reference to Bodenstein's observation that the reaction has a negative temperature coefficient, it was seen that he accounted for this fact by assuming a temporary association of nitric oxide molecules in the form of a complex, the life of which diminishes with rise of temperature The present work indicates that this complex is best formed when the nitric oxide molecules are held close together by adsorption on glass surfaces If these complex molecules are free to evaporate from the surface and possess a short life in the gas phase, then it is clear that the reaction may appear to be homogeneous when subjected to measurements of its kinetics and that it will have a negative temperature coefficient owing to the shortening of the life of the complex with rising temperature. It is seen from the present experiments that these complex molecules are not formed in the gas phase (as was believed by Bodenstein).

The experiments described as "cleaning" experiments showed that the author's observations were obtainable without any recourse to intensive drying. It was also shown that the admission of water to the non-reactive gases had no effect and therefore inhibition of reaction is wholly due to surface conditions of the containing vessels, "drying" in itself having no effect.

The experimental observations of J. W. Smith were shown to be quite correct but his explanation of his observations needs much modification. A new compound has been discovered which, as far as the author can determine, appears to be $2NO.P_2O_5$. The existence of the nitrogen peroxide-phosphoric anhydride compound described by Smith is unconfirmed. *It has been shown that Smith's work does not prove that* intensely dried nitric oxide and oxygen are incapable of reaction.

The electrical conductivity of thin metallic films. 3. Alkali films with the properties of a normal metal.

By A. C. B. Lovell. (*Communicated by A. M. Tyndall, F.R.S.—Received* 4 *February* 1938.)

The technique previously described for measuring the electrical conductivity of thin films deposited by evaporation, in very high vacua, on clean substrates, has been modified to include the deposition of thick films of the alkali metals.

Caesium *films* 10,000 A *thick have resistivities only* 4 % *greater than the* bulk metal, and identical temperature coefficients. Thick rubidium and potassium films have rather higher resistivities, but possess the same temperature coefficients as the bulk metal.

Evidence is produced that the films are polycrystalline, and that the higher resistivities are due to a simple residual resistance The gradation in properties of the thick films is shown to follow consistently from the properties of the very thin films investigated previously.

From these results and other considerations it is concluded that there are no grounds for the belief that thin films differ essentially in structure from the normal metal.

The Zeeman and Paschen-Back effects in strong magnetic fields.

By P. L. Kapitza, F.R.S., P. G Strelkov and E I. Laurman. (*Received* 7 *February* 1938)

1. A method is described for studying the Zeeman and Paschen-Back effects in magnetic fields up to 320,000 gauss.

2. It is shown that the Zeeman splitting is within the limits of experimental error proportional to the magnetic field and obeys the theoretical predictions previously verified only in weaker fields

3. In strong magnetic fields we were unable to discover any displacement of the centre of gravity of the splitting pattern, which again is in agreement with the theoretical prediction

4. The Paschen-Back effect was studied in fields up to 300,000 gauss on the beryllium doublet, and it was shown that the splitting accurately followed the theoretical predictions, and that the intensities of the various components agreed qualitatively with the theory

5 The initial stages of the Paschen-Back effect were also observed for the zinc triplet ^3P–^3S.

The lift and moment on a flat plate in a stream of finite width. By T. H. HAVELOCK, F.R.S. (*Received* 8 *February* 1938.)

The paper gives a new treatment of the problem of a flat plate in a stream bounded by plane parallel walls, including circulation round the plate. The plate is considered as the limiting case of the elliptic cylinder, an integral equation is obtained whose solution by continued approximation leads to expansions for the lift and moment on the plate. The solution is modified to give similar results when the stream is bounded by parallel free surfaces, taking the boundary condition at a free surface in an approximate form; and a further modification gives the case when one boundary of the stream is a plane wall and the other is a free surface. The problem of the elliptic cylinder in general is also considered with reference to the moment of the forces when the stream is bounded by plane walls and when there is no circulation.

Quantum theory of Einstein-Bose particles and nuclear interaction. By N. KEMMER. (*Communicated by S. Chapman, F.R.S.—Received* 9 *February* 1938.)

It is shown that there are four inequivalent but equally simple possibilities of formulating a field theory of Einstein-Bose particles, in which a positive expression for the energy density exists. Any of these formalisms might tentatively be accepted as a description of the "heavy electron" Considerations of relativistic invariance show that two independent expressions for the interaction of these particles with protons and neutrons can be chosen in each of the four cases. Taking account of the interaction terms, the general Hamiltonian form of the theories is stated and the quantization is performed The resulting proton-neutron potential is determined and it is found that its sign and spin-dependence agrees with reality in only one of the four cases, namely in the case based on the equations of Proca (1936). The (divergent) self energies of the proton or neutron resulting from the interaction studied are evaluated

The photosensitivity of diphenylamine-*p*-diazonium sulphate measured by the method of photometric curves. By C. F. GOODEVE and L. J. WOOD. (*Communicated by C. K. Ingold, F.R.S.—Received* 12 *February* 1938.)

The method of photometric curves, used previously to measure the photosensitivity of visual purple solutions, has been applied to the bleaching of diphenylamine-*p*-diazonium sulphate.

The diazonium salt was bleached with light of wave-length 365 $m\mu$. and it was found to have a quantum efficiency of $0\cdot34 \pm 0\cdot02$, independent of concentration and temperature. The photosensitivity was also found to be unaffected by the removal of dissolved oxygen and by the addition of an internal filter.

These results have been compared with those obtained with visual purple, and their photochemical significance discussed.

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

18 MARCH 1938

The control of beating and of micro-fibrillation by means of potassium, calcium and sodium in the chick embryo heart. Potassium fibrillation. By P. D. F. MURRAY (*Communicated by Sir Henry Dale, F.R.S.—Received 7 February* 1938.)

A previous paper showed that fibrillation induced by calcium in the medium depends upon a rise in the ratio calcium at the cell surfaces/potassium in the cell interiors. In the present paper experiments similar in character to those described in the first paper show that fibrillation by potassium in the medium is (1) more readily produced in hearts previously rendered potassium-poor than in hearts which are not potassium-poor, (2) probably more readily produced in hearts previously rendered sodium-poor than in hearts which are not sodium-poor. It is concluded that fibrillation by potassium in the medium occurs (1) when the ratio potassium at the cell surfaces/potassium in the cell interiors is abnormally high, (2) probably also when the ratio potassium at the cell surfaces/sodium in the cell interiors is abnormally high.

The energy loss of penetrating cosmic ray particles in copper. By J. G. WILSON. (*Communicated by P. M. S. Blackett, F.R.S.—Received 10 February* 1938.)

A description is given of measurements of the energy loss of cosmic ray particles in a copper plate, made by means of the cloud chamber method, and the measurements are compared with those previously made in lead.

The results show that the initial reduction of relative energy loss, $(1/E)\,(dE/dx)$, with increasing energy is probably rather less rapid in copper than in lead, but there is no indication that the Bethe-Heitler value for the radiation loss applies to markedly higher energies in the lighter element. For both lead and copper there is a maximum of relative energy loss at $E \sim 1{\cdot}5 \times 10^9$ e-volts. It is shown that this loss is not accompanied by the secondary particles to be expected if it were due to the emission of collision radiation. Further, the value of the relative cross-section in this region varies with the absorber less rapidly than Z^2, and the position of the maximum is almost independent of Z.

Absorption of this type will become of particular importance for elements of low atomic number, and it is suggested that the absorption with a maximum at $E \sim 1{\cdot}5 \times 10^9$ e-volts is responsible for the anomaly in the sea-level energy spectrum of cosmic rays at $E \sim 2{\cdot}5 \times 10^9$ e-volts.

The structure of the walls of parenchyma in *Avena coleoptiles*. By R. D. PRESTON. (*Communicated by W. Stiles, F.R.S.—Received* 14 *February* 1938.)

Consideration of both intact cells and of single walls in oat coleoptiles shows clearly that the cellulose chains of the wall are inclined to the transverse plane, in contradistinction to previously accepted ideas The chains may, therefore, be represented as forming a series of spirals round the cell; and the spiral may be right- or left-handed in different cells, though it retains the same sign throughout in any individual parenchyma. Growth of the coleoptile under constant light conditions, at a temperature of 25° C., appears to involve a change in the inclination of this spiral, in accordance with changes in cell dimension, which may be interpreted in terms of the geometry of the spiral. The ratio of the length of the parenchyma to its girth is shown to be of importance in this respect. If this ratio increases, the spiral becomes steeper; if it decreases, the spiral may be expected to become flatter. It is probable that both turgor forces and active growth of the wall, as well as external conditions, play a part in this change in micellar inclination. Further investigation of coleoptiles, grown under widely different conditions, is clearly necessary before the precise value of such factors can be appreciated

---

Some experimental observations for longitude, made by theodolite, fitted with a shutter eyepiece. By J. DE GRAAFF HUNTER, F.R.S (*Received* 14 *February* 1938.)

Precise astronomical position determination is required in geodesy and geophysics. Longitude observations depend on timing star transits, and the observation errors, both systematic and accidental, are relatively large. At fixed observatories this is mostly overcome by the moving-wire micrometer; but the heavy outfit precludes its wide use in field work, where longitude precision is ordinarily much lower than latitude precision.

The present paper describes a new method, applicable to a theodolite, to enable the triangulator to fix both latitude and longitude with equally high precision; with little additional equipment. The basic principle is to observe the position of a star *at instants controlled by the chronometer*; instead of observing the time at which it reaches definite positions.

A special eyepiece contains a shutter, operated electro-magnetically every third second by the chronometer; operation period being say 0·07 sec. Operation may also be every second, for comparison of time with rhythmic radio signals. Further, a scale with horizontal lines is provided in the eyepiece and rendered luminous; while the shutter is not visible. The scale readings of star position at some twenty successive appearances are recorded. Their mean, converted to arc seconds and applied to the corrected vertical reading, gives the zenith distance at a known precise chronometer instant. With suitably chosen stars, both latitude and longitude are determinate.

Observations over several years, made on the shutter principle with a meridian transit, have shown no personality. Recently, with a Tavistock theodolite of 10 in. focus, the M.S.E. of longitude from a pair of stars = ±0 075 sec. ≡ ±0·7 equatorial sec. of arc; the over-all time of observation of the pair being 8 min. This result is fully as precise and rapid as that of a Talcott latitude pair with a zenith telescope of greater power.

The maternal effects on growth and conformation in Shire horse-Shetland pony crosses. By A. Walton and J. Hammond, F.R.S.

(*Received* 15 *February* 1938.)

1. Reciprocal crosses between the large Shire horse and the small Shetland pony have been made by means of artificial insemination.

2. At birth the foals were approximately proportional in weight to the weights of their mothers and about equal to foals of the pure breeds to which the mothers belonged. The cross-foals from the Shire mare were three times the size of the cross-foals from the Shetland mares. Maternal regulation of foetal growth was very marked and obscured any genetic differences.

3. After weaning, when the foals were under the same nutritive conditions, genetic differences appeared. The foals from the Shire mares grew much less rapidly than pure Shire foals, and the foals from the Shetland mares grew much more rapidly than pure Shetlands. At about 18 months an equilibrium point was reached at which the relative growth rates of the cross-foals and the pure Shetland remained constant. At 3 years the difference between the reciprocal crosses is still marked and is apparently permanent.

4. Differences in the proportions of the animals, when size differences are eliminated, were not so marked as differences in weight and the influence of nutrition not so obvious.

5. The mechanisms, by which maternal regulation may be brought about, are discussed and three possibilities suggested: (*a*) maternal regulation of foetal nutrition, (*b*) maternal hormonal control, (*c*) cytoplasmic inheritance.

6. The bearing of these results on the theoretical concept of growth is discussed.

7. The experiments illustrate the interplay of nutritional and genetical factors which are involved in development.

The biological characters of spontaneous tumours of the mouse, with special reference to rate of growth. By A. Haddon. (*Communicated by E. L. Kennaway, F.R.S.—Received* 15 *February* 1938.)

The following findings were obtained in a study based on 336 spontaneous tumours in the mouse. The occurrence of multiple primary neoplasms approached a chance distribution, although the approximation was statistically inadequate to justify the assumption of actual random incidence. In the great majority of cases the linear measurement of tumour size increased linearly with time, although a small proportion showed exponential increase. No significant association could be established between rate of growth and the incidence of metastasis. A significant correlation was found between tumour size attained and the incidence of metastasis. Time of duration of the primary tumour was found to be a highly important single factor in determining the occurrence of metastasis. No significance could be attributed to a somewhat increased proportion of metastasis in mice with multiple primary tumours. Tumours situated in the caudal half of the body were observed to possess a mean growth rate significantly higher than the mean for similar tumours cephalic in position. No relation was found to exist between location of tumour and production of metastases. When tumours of extreme types were studied, it appeared justifiable to conclude that those of differentiated adenomatous structure more often possessed low rates of growth as compared with the higher rates frequently manifested by dedifferentiated and anaplastic histological types. In the case of pregnant animals no evidence was found to suggest that gestation influenced the rate of tumour growth, but parturition and the onset of lactation were not uncommonly followed by a temporary retardation. Slow growth was frequently observed in the earliest stages of tumour development, such examples attaining their maximal growth rate only after a variable and often considerable interval. Mention is made of the possible etiological significance of this phenomenon. Apart from the effect described as consequent on parturition and lactation, retardation during the later stages of growth was mainly due to incidental factors such as bacterial infection of the tumour substance.

The influence of carcinogenic compounds and related substances on the rate of growth of spontaneous tumours of the mouse. By A Haddon. (*Communicated by E. L. Kennaway, F.R.S.—Received* 15 *February* 1938.)

Parenteral administration of 1 . 2 : 5 . 6-dibenzanthracene in mice bearing spontaneous neoplasms (mainly carcinomata of the mammary gland) resulted in most cases in a prolonged inhibition of the rate of tumour growth. The degree of individual response varied from slight to marked retardation, and a few cases manifested active regression, either partial or complete. The same result, with a corresponding degree of variation, was produced by the carcinogenic substances 1 : 2 : 5 : 6-dibenzacridine, methyl cholanthrene and styryl 430. The non-carcinogenic compounds pyrene and 1 . 2 : 3 : 4-dibenzanthracene provoked either no response or a transient interference with growth rate followed by complete recovery. Instances are given in which the non-carcinogenic substances acenaphthanthracene and 1 : 2 : 5 : 6-dibenzphenazine led to a retardation of growth not different from that produced by carcinogenic compounds. Administration of large doses of oestrone benzoate produced in some cases no inhibition and in others a moderate retardation with a tendency to recovery. The inhibitory response brought about by carcinogenic compounds and by certain related non-carcinogenic compounds is attributed to the possession by such substances of toxicity of a special kind.

The scattering of fast β-particles by mercury nuclei. By F. C. CHAMPION and A. BARBER. (*Communicated by C. D. Ellis, F.R.S.—Received* 17 *February* 1938.)

The scattering of β-particles of about 1 mV energy by mercury nuclei is examined with an expansion chamber for angles of scattering greater than 20°. The absolute number scattered is much less than would be expected from Mott's formula. There is also indication of more inelastic collisions than are predicted by the theory of Bethe and Heitler. Photographs are reproduced which show the sudden stoppage of a fast β-particle with the simultaneous production of what appear to be low energy pairs of positrons and electrons It is suggested that the results may be qualitatively explained on the assumption of a repulsive field between electrons and nuclei for close distances of approach.

A high-temperature Debye-Scherrer camera and its application to the study of the lattices pacing of silver. By W. HUME-ROTHERY, F R.S., and P. W. REYNOLDS. (*Received* 17 *February* 1938.)

A high-temperature Debye-Scherrer camera is described for use at temperatures up to 1000° C. The camera is designed primarily for the accurate determination of lattice spacings of metals and alloys. The specimen is contained in a very thin-walled sealed silica tube in order to prevent change of composition of the specimen owing to volatilization, oxidation, etc The temperature is measured by means of a thermocouple, and can be controlled to within ± 1° C. of the desired value.

The lattice spacing of silver is determined accurately between 20 and 943° C. The values agree to within 0·0001 A with the data of Scheel for the expansion of massive bars of silver up to 500° C. The values of Keesom and Jansen for the coefficient of expansion of silver in the range − 253° C. enable the lattice spacings at low temperatures to be calculated, so that the lattice spacing of silver is now known from 20° above the absolute zero to 18° below the melting-point. It has been found possible to express the results in the form of an equation such that only four constants are required to express the coefficient of expansion over the whole range. This equation is of such a nature that the coefficient of expansion vanishes at the absolute zero in agreement with theoretical requirements. One term in the equation becomes vanishingly small above 0° C., and the coefficient of expansion from room temperature to the melting-point can be expressed in terms of two constants only. The greatest difference between the calculated and observed values is equivalent to 1 part in 10,000 in the lattice spacing. An equation of the same type has been found to hold for other substances.

The dispersion formula for nuclear reactions. By P. L. Kapur and R. Peierls. (*Communicated by R. H. Fowler, F.R.S.—Received* 17 *February* 1938.)

The dispersion formula of Bethe and Placzek, which expresses the cross-section for nuclear reactions in terms of the virtual levels of the compound nucleus and their widths, is derived without using the assumption that the motion of the incident particle within the nucleus can in any approximation be regarded as a one-body problem.

It is possible to derive a formula of the type of the dispersion formula without any assumption at all about the mechanism of the reaction. However, in the general case this formula contains more parameters than the usual dispersion formula and is therefore of less practical value.

The dispersion formula of Bethe and Placzek is shown to hold rigorously if only the widths of all virtual levels which contribute essentially to the expression are much smaller than the distances between the levels.

An X-ray study of the iron-nickel-aluminium ternary equilibrium compound. By A. J. Bradley and A. Taylor. (*Communicated by W. L. Bragg, F.R.S.—Received* 18 *February* 1938.)

X-ray powder photographs of slowly cooled alloys of iron, nickel and aluminium have been used to construct a phase diagram for all compositions. The system falls into two distinct portions. Up to 50 atomic per cent of aluminium there are only body-centred cubic and face-centred cubic structures. Beyond 50 per cent of aluminium, the diagram is extremely complex and will be described in a later paper.

The face-centred cubic phase field includes superlattice structures related to Ni_3Al or Ni_3Fe (α_1), as well as disordered structures (α). It encloses a miscibility gap where two face-centred cubic superlattice phases (α_1 and α_1'') are in equilibrium. As the iron content increases, the gap closes.

The body-centred cubic phase field includes superlattice structures related to Fe_3Al (β_1) and FeAl or NiAl (β_2). There is a two-phase field where the iron-rich β phase (without a superlattice) is in equilibrium with the β_2 phase. In consequence, the system is much more complicated than had previously been supposed. Instead of a two-phase field separating the face-centred cubic and body-centred cubic areas, there are four separate areas, comprising three two-phase fields and a three-phase field. In the latter a face-centred cubic structure (α) is in equilibrium with two body-centred cubic structures (β and β_2). The two-phase fields are $\alpha+\beta$, $\alpha+\beta_2$ and $\beta+\beta_2$ respectively. The corners of the three-phase triangle are placed at the following compositions, α, 62, 31, 7; β, 90, 7, 3; β_2, 28, 41, 31. (The numbers represent atomic per cent iron, nickel and aluminium respectively.)

A note on Guggenheim's theory of strictly-regular binary liquid mixtures.

By G. S. RUSHBROOKE. (*Communicated by R. H. Fowler, F.R.S.—Received 18 February* 1938.)

An attempt is made to improve and extend Guggenheim's tentative theory of strictly regular binary liquid mixtures by using Bethe's indirect method to avoid the unknown combinatory factor in the partition function for the liquid phase. The expression found in this way for the configurational energy of the mixture differs somewhat from that proposed by Guggenheim, and consequently formulae for the thermodynamic functions are also modified.

The conditions under which such solutions will exhibit critical mixing phenomena are also discussed and the variation with temperature of the concentration at which the liquid mixture begins to separate into two phases is investigated.

Though for simplicity the theory is developed on the supposition that the molecules of the liquid are not close-packed, numerical calculations for face-centred cubic packing, included in the last section, suggest that this does not significantly affect the results.

Investigations on Mediterranean Kala-azar. XI. A study of leishmaniasis in Canea (Crete).

By S. ADLER, O. THEODOR and G. WITENBERG. (*Communicated by Sir Henry Dale, F.R.S.—Received 22 February* 1938.)

The epidemiology of leishmaniasis in Canea is discussed. Human visceral leishmaniasis, canine visceral leishmaniasis, and *Phlebotomus major* have an identical distribution in Canea. *P. major* is the only sandfly in Canea of importance for the transmission of visceral leishmaniasis. *P. perniciosus* var. *tobbi* and *P. perfiliewi* are absent, and *P. chinensis* var. *simici* is not common.

The bionomics of *P. major* in Canea are discussed.

The hitherto unknown male of *P. larousseі* is described and the relation of this species to *P. vesuvianus* and *P. canaaniticus* is discussed. *P. vesuvianus* is considered a synonym and *P. canaaniticus* a variety of *P. larousseі*.

The clinical condition of naturally infected dogs is improved and the infection rate in *P. major* fed on these animals reduced by placing them on a diet of fresh meat without any further treatment.

A number of naturally infected dogs have difficulty in barking owing to infiltration of the vocal chords with macrophages and plasma cells.

P. major, *P. chinensis* var. *simici* *P. sergenti* and *P. papatasii* were infected with *Leishmania infantum* by feeding on naturally infected dogs, on Syrian hamsters and a spermophil. The infection rates in *P. sergenti* and *P. papatasii* were very low as compared with *P. major*.

P. sergenti, *P. papatasii* and *P. major* were infected with *L. tropica* by feeding on lesions in human beings in Canea. The infection rate in *P. papatasii* is very low as compared with that in *P. sergenti*. *P. papatasii* plays no significant role in the transmission of *L. tropica* in Canea, and *P. sergenti* appears to be the main vector of *L. tropica* in Canea.

The Canea strain of *L. tropica* is biologically different from the Palestinian one. In contrast to the latter it produces a very low infection rate in *P. papatasii* both in Greece and Palestine.

A human being was infected with *L. tropica* by the inoculation of flagellates from an artificially infected *P. sergenti*.

The general distribution of sandflies of the major group with relation to visceral leishmaniasis in the Old World is discussed.

The accurate analysis of gaseous mixtures.

By B. LAMBERT and D. J. BORGARS. (*Communicated by C N. Hinshelwood, F.R.S.—Received 22 February* 1938.)

An apparatus is described for the accurate analysis of small volumes of gas mixtures.

It has been found possible to obtain results accurate to 0 02 per cent for the determination of the percentage content of one constituent of a gaseous mixture using about 10 c.c. of gas. The apparatus is also suitable for "microanalyses" of gaseous mixtures using 1 to 2 c.c. of gas mixture and results accurate to 0·1 per cent are obtainable.

Difficulties associated with the accurate determination of carbon monoxide in a gaseous mixture are discussed and methods of overcoming them are described.

Atomic rearrangement process in the copper-gold alloy Cu_3Au. II.

By F. W. JONES and C. SYKES. (*Communicated by W. L. Bragg, F.R.S.—Received* 24 *February* 1938.)

The relation between the size of nuclei and electrical resistance has been determined for the copper-gold alloy Cu_3Au in both the annealed and cold worked states. The experimental results are interpreted using the hypothesis that nuclei are anti-phase with narrow boundaries. For large nuclei a linear relation, between resistance and the number of boundaries per unit length, is found from which the reflexion coefficient of a single boundary has been deduced. This reflexion coefficient is the same for both cold worked and annealed material although the rates of growth of the nuclei in the two cases differ considerably.

ABSTRACTS
OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

5 April 1938

On the theory of heavy electrons and nuclear forces. By H. J. Bhabha. (*Communicated by R. H. Fowler, F.R.S.—Received* 28 *February* 1938.)

A theory is developed based on the idea that the proton and neutron are two states of the same particle, which can go over from one state to the other by the emission of a charged particle of mass intermediate between those of the proton and electron, as originally suggested by Yukawa. These "U-particles" are described by four wave functions. Quantization of the theory leads as usual to positive and negative U-particles with a spin of one unit. The U-particles are identified with the heavy electrons of cosmic radiation. The theory leads uniquely to short range forces of the Heisenberg and Majorana type of such a sort as to allow one to make the ground state of the deuteron the triplet state, *the sign of the Majorana force being not at our choice*. The range of the forces is connected with the mass of the U-particles as before, and demands a mass of some two hundred times the electron mass. The relativistic generalizations of a pure Heisenberg and pure Majorana force are given. The relativistic scattering of U-particles by protons and neutrons is calculated. The theory also leads to showers of Heisenberg's type, but consisting of heavy electrons and heavy particles only.

The statistical mechanics of condensing systems. By M. Born and K. Fuchs. (*Communicated by E. T. Whittaker, F.R.S.—Received* 1 *March* 1938.)

J. E. Mayer's general theory of gases was the subject of a discussion at the Van der Waals Centenary conference in Amsterdam (1937) where objections were raised against his explanation of condensation. We present a new form of Mayer's theory using the method of complex integration by which it is possible to show rigorously that Mayer's statements are essentially correct.

The starch-iodine coloration as an index of differential degradation by the amylases.

By C. S. Hanes and M. Cattle. (*Communicated by F. F. Blackman, F.R.S.—Received* 7 *March* 1938.)

The alterations in iodine coloration have been followed by a spectrophotometric method during the action of five representative amylases on starch. It is found that the so-called saccharogenic and dextrinogenic types of amylase are distinguishable from the early stages of their action by the manner in which the absorption characteristics of the iodine compounds are changed The relation between the alterations in the iodine colouring property and the liberation of reducing groups throughout the different degradative processes has been examined; such *colour/reducing power* relationships are shown to constitute a diagnostic criterion for comparing the action of different amylases.

The problem of the constitutional basis of the iodine colouring property of starch and starch dextrins is considered in the light of these quantitative data. It would appear that the hue of iodine colour exhibited by a particular product is to a large extent independent of the length of its chain molecule but is governed mainly by some factor such as the mode or degree of association of the chemical units as they exist in molecular aggregations. This hypothesis is in harmony with the available evidence from purely chemical studies, on the one hand, and with conceptions of the mode of attack of the starch molecule by the different enzymes, on the other.

The dielectric polarization of a *n*-long chain ketone at constant volume and variable temperature.

By A. Müller. (*Communicated by Sir William Bragg, P.R.S.—Received* 7 *March* 1938.)

The present work is a continuation of a previous investigation on the dielectric polarization of a long chain ketone. The substance, instead of being allowed to expand at atmospheric pressure, is now kept at constant volume. This constraint produces a very marked change in the behaviour of the substance when compared with previous results. It is shown that the rise in polarization at constant pressure goes parallel with the large expansion which precedes the melting.

With regard to the theory of cooperative phenomena the present experiments show the effect of a variation of the boundary conditions upon the behaviour of the substance.

Transference of induced food habit from parent to offspring, III.

By D. E. Sladden and H. R. Hewer. (*Communicated by E W. MacBride, F.R.S.—Received* 10 *March* 1938.)

A modification of food habit has been induced in stick insects and has been transmitted through a number of parthenogenetic generations. The anomalies of the earlier communications have been satisfactorily explained.

Formation of negative ions at metal surfaces. By R. A. Smith. (*Communicated by N. F. Mott, F.R.S.—Received* 10 *March* 1938.)

Various processes of formation of negative ions at metal surfaces, and the conditions under which they may take place are discussed. Detailed calculations are given for the conversion of Hg^+ ions into Hg ions at a nickel surface. The positive ions are assumed to be neutralized by capturing an electron from the metal, and to form excited atoms which subsequently capture another electron in falling into their ground state. This process is shown to account for the experimental results of Arnot and Milligan. The calculated probability of formation by this process is $1 \cdot 4 \times 10^{-3}$ for 200 V Hg^+ ions striking a nickel surface normally. The process of simultaneous capture of two electrons from the metal is shown to be improbable.

A negative ion formed near a metal surface for which the work function is greater than the electron affinity of the corresponding atom or molecule will have a considerable chance of being neutralized through one of its electrons passing into an unoccupied level in the metal. It follows that slow positive ions or metastable atoms will be ineffective as a source of negative ions.

The formation of atomic negative ions from molecular positive ions is discussed. Calculations for the formation of H^- ions from H_2^+ ions and from protons are given. The value obtained for the probability of formation of H ions from 140 V protons striking a surface with work function less than 3·37 e-volts is about 4×10^{-3}. It is shown that H^- ions will not be formed from protons at a surface with work function greater than 3·37 e-volts unless simultaneous capture of two electrons takes place. For formation of H^- ions from H_2^+ the observed probability is only $1 \cdot 04 \times 10^{-5}$ for 200 V H_2^+ ions striking a nickel surface. This low value is shown to be due to the small probability of formation of the intermediate excited H atoms by dissociation of the incident ions. Clearly, further experiments, using separated beams of protons and H_2^+ ions, are required in order to clarify the processes taking place when atomic negative ions are formed from molecular positive ions

---

Showers produced by the penetrating cosmic radiation. By W. Heitler. (*Communicated by N. F. Mott, F.R.S.—Received* 7 *March* 1938.)

The theory of the heavy electron developed by Fröhlich, Heitler and Kemmer, has been applied to the passage of fast heavy electrons through matter. It is shown that various types of showers are produced (Y = heavy electron, P = proton, N = neutron).

(1) $Y^+ + N = P + \hbar\nu$. The light quantum produces an ordinary cascade shower.

(2) Multiple processes of the type $Y^+ + N = P + Y^+ + Y^-$ and of higher order are possible. They lead to penetrating showers produced by penetrating particles.

(3) In a heavy nucleus: $Y^+ + N = P$. The energy of Y^+ is restored in the nucleus which subsequently evaporates emitting a few protons and neutrons (proton shower) accompanied by electrons and possibly also heavy electrons.

The inverse process to (1) leads to the creation of heavy electrons by light quanta. It is shown that the order of magnitude of the cross-section is sufficient to explain all heavy electrons at sea-level as secondaries produced in the high atmosphere.

Owing to the breakdown of the theory for energies $\gg \mu c^2$ ($\gg 10^8$ e-volts) all these processes can only be discussed qualitatively and their cross-section can only be calculated approximately for energies of the order 10^8 e-volts or less. For this energy the cross-sections have a reasonable order of magnitude.

The formation of helium molecules. By F. L. Arnot, Ph.D. and Marjorie B. M'Ewen, B.Sc. (*Communicated by H. S. Allen, F.R.S.—Received 11 March 1938.*)

An investigation of the formation of ionized molecules in argon, neon and helium has been made by the balanced space-charge method which had previously been used by the authors to study the formation of mercury molecules. No evidence of molecular ionization was found in argon and neon. In helium molecular ionization sets in at the resonance potential. The results show that these molecular ions are formed by the attachment of metastable atoms to normal atoms according to the process

$$\mathrm{He}(1s^2,\ {}^1S_0) + \mathrm{He}'(1s\,2s,\ {}^3S) \rightarrow \mathrm{He}_2^+(1s\sigma^2\,2p\sigma,\ {}^2\Sigma_u) + \epsilon.$$

The appearance potential of the molecular ions is the energy of the $2\,{}^3S$ state, 19·77 V, which is 1·19 V greater than the ionization potential of the helium molecule.

Attention is drawn to the fact that these molecular ions are formed from excited atoms in S-states, whereas we have shown that mercury molecular ions are apparently formed only by excited atoms in P-states.

An accurate determination of the range-distribution curve of the radioactive alpha particles from ^{8}Li. By C. L. Smith and W. Y. Chang. (*Communicated by J. D. Cockcroft, F.R.S.—Received 12 March 1938.*)

The paper describes a method for determining the number-range distribution curve for the α particles from radioactive lithium. These α particles are shown to consist of a continuous distribution extending up to a maximum of $6{\cdot}95 \pm 0{\cdot}1$ cm. ($7{\cdot}75 \pm 0{\cdot}05$ mV). The shortest range observed was 0·65 cm. (1·2 mV) and in the region from 0·65 cm. up to 6·6 cm. there is a linear relation between the logarithm of the number of α particles and their range Comparison is made with the distribution curves obtained by other workers and it is also shown that the upper limit of the energy of the α particles explains the failure to observe the protons emitted in the reaction:

$${}^7\mathrm{Li} + {}^2\mathrm{D} = {}^8\mathrm{Li} + {}^1\mathrm{H}.$$

Upper and lower limits for the mass of ^{8}Li are calculated from the data and shown to be respectively 8·0253 and 8·0246.

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

27 April 1938

The structure and relations of the human premaxilla. By E. H. Johnson.

(*Communicated by F Wood Jones, F.R S —Received* 10 *March* 1938)

The structure and relations of the human premaxilla have been described. It has been shown to be vestigial in character compared with the homologous bones in lower animals.

The premaxillae are completely covered on the facial aspect by the incisor processes of the maxillae. "Suture lines" on the facial surface have been shown to be due to superficial fissures between the ridges of developing bone on the frontal process of the maxilla. From the conditions observed in the region of the floor of the nasal cavity, it seems that the double narial margins which occur in prognathous races are attributable to an incomplete passage of the incisor processes of the maxillae towards the anterior nasal spine.

The anterior alveolar walls of the incisor tooth are maxillary.

The ascending process of the premaxilla becomes entirely restricted to the internal nasal surface of the maxilla, because of relative differences in the rate of bone growth in this region. In cleft palate, the variation in position of the cleft is not due to a division through the premaxilla, but to the lateral incisor and its alveolus becoming displaced to the lateral side of the cleft. Pathological "exfoliation of the premaxilla" is not indicative of its structural form.

The premaxilla in man has been compared with that in other Primates and has been shown to differ in form, relations, and the time of closure of the sutures. The peculiar shutting off of the premaxilla on the face by the incisor processes of the maxillae is to be regarded as a specific character of man.

The distribution of carbon dioxide in the hen's egg. By J. Brooks and J. Pace. (*Communicated by Sir Joseph Barcroft, F.R.S.—Received* 14 *March* 1938.)

The combined CO_2 of egg-white in equilibrium with different partial pressures of CO_2 was measured over the range $p = 0{\cdot}05$–$1{\cdot}0$ atm. CO_2 at 25° C. An increase in CO_2 pressure increased the concentration of combined CO_2. It was assumed that the combined CO_2 was present in the form of bicarbonate. On this basis the buffer value per g. protein was $4{\cdot}8 \times 10^{-5}$ for the pH range 6·6–7·8. The reasonable values calculated for the pH of the white and the buffer value of the proteins supports the assumption that the amount of carbamino—CO_2 present is small.

The value of the Bunsen solubility coefficient for CO_2 and white at 25° C. was 0·71. This value is about 8 % greater than the value calculated from the salt and water content of the white. It is suggested that, as in blood serum, the excess solubility is caused by the presence of traces of lipoids.

The uptake of CO_2 by shell and yolk was measured. The retention of CO_2 and buffer values of egg-white and the relation between the salts of shell and white have been discussed.

Radioactive nodules from Devonshire. By M. Perutz. (*Communicated by J. D. Bernal, F.R.S.—Received* 15 *March* 1938.)

Globular concretions surrounded by bleached haloes are found in Permian Beds west of Budleigh Salterton, Devonshire. Sections show that their internal structure is complex, a sandy matrix being impregnated by a black hard material in radial and concentric shoots. Contact photographs of a plane section show a distribution of radioactive material closely connected with the structure of the black impregnation. From measurements of the intensity of the radioactivity on an ionization counter a content of 0·4 % uranium was calculated and later confirmed by chemical analysis.

The physical properties of the black part are caused by an apparently amorphous impregnation of vanadium oxides, various other ores also being present in smaller concentrations. Three classes of nodules are described and the possibilities of their origin are discussed.

The last part contains a chemical and geometrical description of the bleached haloes, and after discussion of explanations of their formation suggested by earlier authors the ionizing effect of radon is proposed as an agent which might be partly responsible for the bleaching.

Application of reciprocity to nuclei. By M. Born. (*Communicated by E. T. Whittaker, F.R.S.—Received* 15 *March* 1938.)

The formula for the distribution of quantum states, which follows from the principle of reciprocity and the assumption of a closed p-space, is applied to some properties of nuclei. The results can be considered as confirmations of the hypothesis. The mass of a particle moving with velocity of light and maximum momentum is of the same order as that of the particles which Yukawa has introduced.

Classical theory of radiating electrons. By P. A. M. Dirac, F.R.S. (*Received* 15 *March* 1938.)

The object of the paper is to set up in the classical theory a self-consistent scheme of equations which can be used to calculate all the results that can be obtained from experiment about the interaction of electrons and radiation. The electron is treated as a point charge and the difficulties of the infinite Coulomb energy are avoided by a procedure of direct omission or subtraction of unwanted terms, somewhat similar to what has been used in the theory of the positron. The equations obtained are of the the same form as those already in current use, but in their physical interpretation the finite size of the electron reappears in a new sense, the interior of the electron being a region of space through which signals can be transmitted faster than light.

Self-consistent field with exchange for calcium and argon. By D. R. Hartree, F.R.S. and W. Hartree. (*Received* 16 *March* 1938.)

The differences between the atomic wave functions calculated by the method of the self-consistent field with and without exchange can in certain cases be plotted in such a way that the results for different atoms fall on approximately the same curve, which then can be used for interpolating between different atoms.

Wave functions so estimated using the results already calculated for Cl^- and Ca^{++} have been used as the basis of calculations of the self-consistent field for K^+ and Ar. For K^+ all the (nl) wave functions have been calculated; the results showed that for Ar all but the outer $[(3s)$ and $(3p)]$ wave functions could be interpolated to adequate accuracy by the method mentioned.

Results are given, and values of the diamagnetic susceptibilities calculated and compared with observed values.

Studies of the post-glacial history of British vegetation. By H. Godwin and M. H. Clifford. (*Communicated by Sir Albert Seward, For.Sec.R.S.—Received* 16 *March* 1938.)

The plant remains of the fen deposits have been analysed, and stratigraphy has been determined from profiles and from extensive borings: the methods of pollen analysis have been employed to indicate the drift of local vegetation phases on the fens themselves.

Part I deals with the Woodwalton Fen area, a part of the fenland margin south of Peterborough, where the fen peats have been little damaged by drainage and peat cutting. It is shown that on the landward side there is a single peat bed, which is separated into two not far from the fen margin, by the tapering edge of a bed of fen clay shown by foraminiferal and diatom analysis to have been laid down in brackish water. This clay, which must have represented a marine transgression, interrupted a phase of extensive development of fen woods, at first mainly alder-oak, and later pine-birch. Tree remains are very abundant. The peat above the fen clay showed clear evidence of the development of acidic sphagnum peat of the kind found only in raised bogs ("Hochmoore"). This type of peat was previously unrecognized in the fens, and it reflects conditions of freedom from flooding by alkaline water. It is probable that this phase corresponds with the Bronze Age, and that in the succeeding Iron Age conditions changed sharply. The acid sphagnum peat is overlaid by the calcareous lake marl of Ugg Mere and Trundle Mere, lakes which were only recently drained. They are thus shown to have had a very recent origin.

Part II of the paper extends the observations in the Woodwalton area towards the sea and towards the southern half of the fens. A series of long sections has been constructed which converge towards Wisbech (upon the main estuary of the last phase of fen history). These sections show that the upper and lower peats, separated by the fen clay, occur regularly and continuously over the entire area. On the seaward side they are overlaid by an upper layer of semi-marine silt deposited in the Romano-British period. By a long series of shallow bores the above series of sections was tied to the known profiles at Wood Fen, Ely. Thence it was clear that the phases of lower peat, fen clay, upper peat and upper silt could be accepted as broad major divisions across the whole of the southern part of the Fenland.

Peat formation in the Boreal period was restricted to places of local wetness, such as deep river valleys. A dry phase at the Boreal-Atlantic transition corresponded with a late Tardenoisian culture horizon. Peat formation became general in the fens in the Atlantic period, and the fens became wooded in the Neolithic period, during the end of which time, or just after which, there was an extensive but shallow marine transgression which caused the fen clay to be formed. The succeeding period in the fens began with "A" Beaker culture and during the Bronze Age they were dry: the fens were either wooded or formed raised-bogs, and were fairly habitable. The ensuing Iron Age must have been wet. The Roman period was marked by the deposition of considerable thicknesses of semi-marine silt in a wide belt on the seaward side of the fens, and in tongues along the courses of the estuaries. There were human settlements upon the silt whilst it was forming, and its present surface shows the remains of dense occupation. The great meres of the Fenland probably formed either in the Iron Age or the Romano-British period.

On the theory of scattering of light. By Hans Mueller. (*Communicated by R. H. Fowler, F.R.S.—Received 16 March 1938.*)

The Krishnan effect can be explained if the fluctuations of the optical anisotropy in the medium are not independent in neighbouring volume elements. By applying Brillouin's method to longitudinal and transversal waves the Krishnan effect is calculated for arbitrary directions of observation. It is found that the reciprocity relation is always valid. Every substance which has a Krishnan effect should show a Mie effect. Both effects are related to the photoelastic properties in the case of solids, and to the constants of streaming birefringence in the case of liquids.

The Krishnan and Mie effects must occur for temperature scattering in solids. In isotropic solids the scattered light must consist of two pairs of Doppler components, for crystals three different Doppler shifts can be expected. The Krishnan effect can occur in ordinary liquids if Lucas' transversal waves give a noticeable contribution to the scattering. For temperature scattering the depolarization $\rho_h = V_h/H_h$ is always larger than 1.

A new theory of the scattering of light in glasses is proposed. It is based on the assumption that glasses contain a random distribution of "frozen in" strains. Slightly below the solidification temperature these strains are normal ones, but at lower temperatures shearing strains are created due to temperature contraction. ρ_h is smaller than 1 because the shearing strains are always smaller than the normal strains. It is found that Krishnan's data are in excellent agreement with those calculated for glasses for which the photoelastic constants are known.

Liquids with molecular clusters can be treated as if they were uniform liquids with a distribution of internal strains. They show $\rho_h < 1$ because the strains are predominantly normal ones. A relation between the Mie effect for critical opalescence and the deviation from the λ^{-4} law is found to agree with Rousset's observation.

It is pointed out that scattering data furnish a Fourier analysis of the optical variations within the medium and can be used to determine the size of molecular clusters and of colloidal particles.

Gans' theory of scattering by molecular clusters is discussed and it is shown that it agrees with Krishnan's reciprocity relation.

The two-stage auto-ignition by hydrocarbons. By G. P. KANE. (*Communicated by A. C. G. Egerton, F.R.S.—Received* 17 *March* 1938.)

Previous investigations into the spontaneous ignition under pressure of the higher paraffins and olefins containing more than three carbon atoms have shown that in the temperature range between *ca.* 270–400° C, ignition occurs by a two-stage process preceded by an induction lag t_1 before the formation of a cool flame and a second lag t_2 before the subsequent ignition of the cool flame products; increasing pressure shortens both these lags, and although kinetic relationships have been developed it has not been possible adequately to test them owing to the extreme violence of the ignitions at pressures much above the minimum ignition pressure.

An optical recording manometer is described whereby it has been possible to measure t_1 and t_2 at pressures up to 15 atm. with an accuracy of 1/100th of a second.

With propane, t_2 decreases more rapidly than t_1 with increasing initial pressure and at a critical pressure (about 8 atm.) the two-stage is replaced by a single-stage process; the induction lag then decreases very rapidly with pressure with propylene, where the induction lags are much greater, no such transition had occurred at pressures up to 12 atm.

The bearing of these results both on the nature of the kinetic processes operative and on the problem of "knock", is briefly discussed.

The morphology of the brachial plexus, with a note on the pectoral muscle, and the twist of its tendon. By W. HARRIS. (*Communicated by W. Trotter, F.R.S.—Received* 17 *March* 1938.)

The primary or ventral divisions of the spinal nerves subdivide again into dorsal and ventral nerves to supply the dorsal and ventral muscles and skin of the limb, dorsal joining dorsal and ventral joining ventral divisions.

The simplest type of plexus consists of two cords, dorsal and ventral, formed by the union of the dorsal and ventral subdivisions of all the primary nerves entering the plexus, as in frogs and toads, and in birds. In mammals the original ventral cord, and to a less extent the dorsal cord becomes split up into different nerve trunks, though dorsal and ventral nerves remain apparently strictly apart.

Examination of the reptilian type of plexus in salamanders, turtles, lizards, crocodiles and the Monotremata demonstrates the passage of dorsal fibres into the median and ulnar branches of the ventral cord, probably for convenience of carriage, facts which suggest that essentially dorsal fibres in Man must enter the ventral trunks early, and thus explain the dorsal cutaneous and muscular supply in the hand by the median and ulnar nerves in Man.

The marsupial plexus is a primitive mammalian type, and varies little with the structure of the animal.

Short-necked mammals like the manatee and the Cetacea have five to seven nerves in the plexus, and long-necked forms like the llama and horse have a compressed plexus.

The Carnivora and quadrupedal Ungulata have lost the fifth cervical nerve from the plexus.

The second dorsal nerve may be considered to be a disappearing feature in the mammalian plexus.

In the Primates with arboreal life and increasing use of the deltoid, the fifth cervical nerve reappears and the plexus becomes increasingly prefixed up to the anthropoids, and the pectoral muscle develops a twist of 180° in the tendon of its lower half, aiding the power of climbing.

The nuclear magnetic moment of copper. By S. TOLANSKY and G. O. FORESTER. (*Communicated by P. M. S. Blackett, F.R.S.—Received* 18 *March* 1938.)

The doublet hyperfine structures of the resonance lines of the copper spectrum, $\lambda\,3247$ and $\lambda\,3274$, have been measured with a quartz Lummer plate. The lines are produced free from reversal effects, the doublet separations being respectively 379 and 405×10^{-3} cm.$^{-1}$. The following hyperfine structure interval factors are calculated. $3d^{10}\,4s\,1\,{}^2S_{\frac{1}{2}} = 197\ 5$, $3d^{10}\,4p\,2\,{}^2P_{\frac{1}{2}} = 14$ and $3d^{10}\,4p\,2\,{}^2P_{\frac{3}{2}} = 4{\cdot}8$ (all in cm.$^{-1} \times 10^{-3}$). The mean nuclear magnetic moment for the two copper isotopes, 63 and 65, is derived from the ground state, $3d^{10}\,4s\,1\,{}^2S_{\frac{1}{2}}$. The value found is $\mu = 2{\cdot}47$ nuclear magnetons, this being probably a better estimate than that given by other terms, since the ground state is spherically symmetrical and thus not affected by quadrupole moment of the nucleus. By adopting Schüler and Schmidt's value for the ratio of the magnetic moments of the two isotopes, it is found that $\mu\,\mathrm{Cu}_{63} = 2{\cdot}43$ and $\mu\,\mathrm{Cu}_{65} = 2{\cdot}54$ nuclear magnetons.

The disintegration of boron by slow neutrons. By C. O'Ceallaigh and W. T. Davies. (*Communicated by E. V. Appleton, F.R.S.—Received* 18 *March* 1938.)

The disintegration of boron by slow neutrons— ${}_5\mathrm{B}^{10}+{}_0n^1={}_3\mathrm{Li}^7+{}_2\mathrm{He}^4+\mathrm{Q}_1$—has been studied in an expansion chamber. Evidence has been found for the emission of three groups of heavy particles of ranges 4·25 ± 0·2 mm., 7·15 ± 0·25 mm. and 8·9 ± 0·4 mm. in air at 15° C. and 760 mm. The 4·25 mm. group is clearly due to the lithium nucleus. On the basis of the recent range-energy determinations of Blewett and Blewett, the 7·15 and 8·9 mm. groups, which must consist of α-particles, would lead respectively to energy releases of 2·45 and about 3×10^6 e-volts. Substitution in the equation of reaction of the masses at present accepted leads to a value for Q_1 of $2{\cdot}99 \times 10^6$ e-volts. The 8·9 mm. α-particles would correspond, therefore, to the production of the Li^7 nucleus in the ground state, and the 7·15 mm. particles to an excited nucleus having an energy of $0{\cdot}55 \pm 0\ 15 \times 10^6$ e-volts above the ground state

Evidence based on the emission of two groups of protons in the reaction

$${}_3\mathrm{Li}^6+{}_1\mathrm{D}^2={}_3\mathrm{Li}^7+{}_1\mathrm{H}^1$$

leads to the prediction of one low-lying excitation level of Li of $0\ 44 \times 10^6$ e-volts. This is confirmed by absorption measurements on the γ-radiation accompanying this disintegration. Our value $0{\cdot}55 \pm 0{\cdot}15 \times 10^6$ e-volts is in satisfactory agreement with the above figure. A study of the γ-radiation arising from the non-capture excitation of Li^7 by α-particles would seem to indicate the presence of two low-lying excitation levels of 0·6 and $0{\cdot}4 \times 10^6$ e-volts, but it is possible that transitions to the $0{\cdot}6 \times 10^6$ e-volts state are not allowed in the above reactions.

The heteropycnosis of sex-chromosomes and its interpretation in terms of spiral structure. By M. J. D. White. (*Communicated by J. B. S. Haldane, F.R.S.—Received* 22 *March* 1938.)

There appear to be two kinds of heteropycnosis met with in the chromosomes of the *Orthoptera saltatoria*, which may be called the reversible and non-reversible types. In the reversible type the same chromosome may show both positive and negative heteropycnosis at different stages. In the non-reversible type only positive heteropycnosis is seen. During negative heteropycnosis the chromosome does not undergo as much thickening as in the "control" autosome. In positive heteropycnosis the chromosome undergoes more thickening.

Both positively and negatively heteropycnotic chromosomes have a spiral structure at metaphase. The number of gyres in the spiral per unit length is believed to be inversely proportional to the diameter of the chromosome, so that it will be greater in negatively heteropycnotic chromosomes than in positively heteropycnotic ones. The direction of coiling of the spiral is not constant in the X-chromosomes of the Tettigonidae, i.e. the same chromosome may show either right-handed or left-handed coiling. In the spermatogonial divisions the direction of coiling may be reversed at the spindle attachment of the chromosome or elsewhere.

A new and anomalous type of meiosis in a mantid.

By M. J. D. WHITE. (*Communicated by J. B. S. Haldane, F.R.S.—Received* 22 *March* 1938.)

The meiosis of *Callimantis* takes place in a manner which is entirely unique. No chiasmata are formed and the first meiotic division is consequently reductional for all parts of the chromosomes. The usual diplotene and diakinesis stages are completely absent. Unlike other mantids hitherto studied this species (*C. antillarum* Sauss.) is *XO* in the male.

The problem of *n* bodies in general relativity theory.

By Sir ARTHUR EDDINGTON, F.R.S. and G. L. CLARK. (*Received* 22 *March* 1938.)

The motion of a system of two bodies (e.g. a double star) is investigated as far as terms of the second order in the potentials. Contrary to a result obtained by Levi-Civita in 1936, we find no secular acceleration of the centre of mass of the system.

The work revealed an error in the standard formulae for the line-element of a system of n bodies, given by de Sitter in 1936. As a result of correcting the error, the equivalent mass M of the system is found to be

$$M = E + \tfrac{1}{2} d^2C/dt^2,$$

where E is the energy (including potential energy) and C the moment of inertia about the centre of mass.

The secretion of crystalloids and protein material by the pancreas.

By S. A. KUMAROV, G. O. LANGSTROTH and D. R. MCRAE. (*Communicated by J. S. Foster, F.R.S.—Received* 25 *March* 1938.)

The concentrations of Na, K, Ca, Cl, HCO_3, and protein and non-protein nitrogen, were determined in series of samples of pancreatic juice secreted by dogs in response to (*a*) constant rate of administration of secretin, (*b*) varied rate of administration of secretin, and (*c*) interrupted administration of secretin (rest period, 2 hr). The use of quantitative spectroscopic methods of analysis in the determination of the metals, and of absorption spectrum methods in the study of the protein composition of the secretion, were important features of the technique.

The fact that the observed concentrations of metals in the secretion are independent of the degree of activity of the gland, is considered to indicate that the glandular membranes offer little resistance to the passage of simple inorganic ions. The *apparent* marked differences in permeability to different ions, as indicated by Ball's (1930) injection experiments, are probably due to a transformation of part of the injected substances to "non-diffusible" forms in the blood stream.

The observed increase in the bicarbonate concentration with increasing rate of secretin administration, and the compensatory relation with the chloride concentration, is interpreted on the basis of (*a*) the formation of at least a part of the bicarbonate as a product of metabolism within the glandular cells, and (*b*) the action of membrane forces, probably of an electrical nature.

A correlation of absorption spectrum measurements with protein nitrogen data admits of two alternative interpretations; (*a*) only one type of absorbing protein is present in the secretion, and a part of all of this serves as a carrier for enzymatically active groups, or (*b*) several types of absorbing protein are present, and these, whether enzymatically active or not, are always secreted in constant proportions.

The interpretation of protein nitrogen and certain other data is given in the accompanying paper.

The processes of synthesis and secretion of protein material in the pancreas.

By S. A. Kumarov, G. O. Langstroth and D. R. McRae. *(Communicated by J. S. Foster, F.R.S.—Received 25 March 1938.)*

A theory of the process of synthesis of protein material by the pancreas, and its secretion in response to secretin administration, is developed The theory leads to expressions which describe quantitatively the behaviour of the protein output in samples of secretion obtained under widely varied experimental conditions. It permits some insight into the fundamental nature of certain of the glandular processes, as well as the calculation of various factors not directly observable in critical experiments. It is rich in suggestions for new problems in connexion with the secretory processes.

The applicability of the Gibbs' adsorption theorem to solutions whose surface tension curves exhibit minima or horizontal portions.

By J. W. McBain, F.R.S. and G. F. Mills. *(Received 25 March 1938)*

The classical Gibbs' adsorption theorem appears to be a limiting law applying to cases where mutual repulsion or oriented dipoles upon the surface and the effects of submerged double layer and of electrification do not also condition the surface tension.

All aqueous and ionizing systems involve the important factor of a submerged double layer of greater or lesser development, powerfully affecting the surface tension, appreciable even for insoluble oil films on water or for pure water itself.

Abundant examples of "Type 3" surface tension curves have been obtained by many authors by every experimental method. In these, the static or true equilibrium surface tension is very greatly lowered in extreme dilution, thereafter remaining constant or passing through one or more maxima or minima, although the dynamic or immediate surface tension is several times greater, nearly that of water. Application of the classical Gibbs' theorem in its exact form to these cases yields a series of obviously impossible results, no matter in what form they are calculated. The different reasons adduced by many authors for avoiding such applications are seen to be gravely at variance with the foundations of chemical thermodynamics.

The factors of orientation and submerged double layer omitted in the Gibbs' formulation are indicated.

It follows from the Gibbs' equation that the surface of solutions of soluble substances is very deep as compared with molecular dimensions, in contrast to the familiar conception of insoluble films on water.

The first spark spectrum of platinum. By A. G. SHENSTONE. (*Communicated by R. H. Fowler, F.R.S.—Received* 25 *March* 1938.)

The paper reports observations on the Spectrum Pt II and classifies a large number of levels. A list of lines is also given including (i) all identified lines between $\lambda 976$ and $\lambda 1242$, (ii) all certain low transition lines and all identified high transition lines between $\lambda 1242$ and $\lambda 4514$.

The adsorption of vapours at plane surfaces of mica. Part II. Heats of adsorption and the structure of multimolecular films. By D. H. BANGHAM and S. MOSALLAM. (*Communicated by D. L. Chapman, F.R.S.—Received* 25 *March* 1938.)

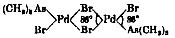
The results are given of measurements of the quantities of benzene, methyl alcohol, and carbon tetrachloride adsorbed at a known surface area of mica at pressures approaching saturation. The adsorption energy is constant whilst the first molecular layer is being formed, but its completion is marked by an abrupt decrease to a value near the normal heat of liquefaction. The graphs of apparent film thickness against relative pressure for the three substances are not widely divergent over much of the range investigated, but the isotherms show discontinuities or sharp changes of curvature at adsorption values which are thought to be related to molecular spacings in the bulk condensed phases. Approximate values are given for the surface tension lowering of the mica, caused by the saturated vapours.

The swelling of charcoal. Part V. The saturation and immersion expansions and the heat of wetting. By D. H. BANGHAM and R. I. RAZOUK. (*Communicated by D. L. Chapman, F.R.S.—Received* 25 *March* 1938.)

The percentage linear expansion x_s of a rod of inactive wood charcoal when exposed to the exactly saturated vapour of methyl alcohol is found to be less than the expansion x_L produced by immersion in the liquid. The expansion is proportional to the surface energy lowering of the charcoal, and this relation is confirmed by comparing $\left(x_L - T\frac{dx_L}{dT}\right)$ with the measured heat of wetting. On immersion of the air-free charcoal the liquid wets all, or nearly all, of the adsorbing surface. From well-supported estimates, previously given, of the specific surface of the charcoal, it becomes possible to assign values to the free and total energy changes per cm.^{2} which accompany saturation and immersion.

The crystal structure of certain bridged palladium compounds. By A. F. Wells. (*Communicated by J. D. Bernal, F.R.S.—Received* 26 *March* 1938.)

The crystal structures of the bridged compounds $[(CH_3)_3As]_2Pd_2Cl_4$ and its bromine analogue have been determined. The crystals of the tetrachloride from alcohol and from dioxane and those of the tetrabromide from dioxane are all isomorphous. A probable structure deduced from the optical properties, cell dimensions and symmetry was confirmed by Fourier projections using visually estimated intensities. The structure of the tetrabromide was investigated in detail, the configuration of the molecule being finally obtained from a section of the three-dimensional electron density distribution. The bridged molecule is found to be planar with the configuration



The interatomic distances are: Pd–Br 2·45 and Pd–As 2·50 ± 0·05 A. The molecules are stacked in columns along 4_2 axes in the space group I 4/*m*, and the columns of molecules are held together by very weak van der Waals forces. The packing of the molecules in this way leaves tunnels through the structure. These remain empty in certain circumstances, as, for example, when the tetrachloride crystallizes from alcohol. On crystallizing from dioxane, however, this compound takes up dioxane of crystallization presumably owing to the fact that the diameter of the dioxane molecules is approximately the same as that of the holes in the structure. The introduction of the solvent is accompanied by a small increase in the cell size and a slight reorientation of the molecules of the palladium compound.

The paramagnetic magneton numbers of the ferromagnetic elements. By W. Sucksmith and R. R. Pearce. (*Communicated by A. M. Tyndall, F.R.S.—Received* 28 *March* 1938.)

A method is described for measuring magnetic susceptibilities in a controlled atmosphere or *in vacuo* at temperatures up to 1500° C. The special precautions required to prevent solid diffusion and minimize evaporation in measurements on metals are discussed. Accurate measurements on the susceptibilities of the ferromagnetic elements, hitherto confined to nickel, are extended to cobalt and iron. The magneton numbers for both these elements are deduced to an estimated accuracy of 1–2 % from the experimental data. It is shown that the existing theory is inadequate to explain the new results obtained.

Electromagnetic induction in non-uniform conductors, and the determination of the electrical conductivity of the earth from terrestrial magnetic variations. By A. T. Price and B. N. Lahiri. (*Communicated by S. Chapman, F.R.S.—Received* 28 *March* 1938.)

The results of previous investigations by Chapman and Price of the induced fields and current distributions, associated with the magnetic daily and storm-time variations, suggest that more precise information as to the distribution of electrical conductivity (κ) within the earth, might be obtained by considering electromagnetic induction in a non uniform sphere The general theory for any non uniform conductor is here considered, and the formal solution for any conductor with spherical symmetry is obtained Detailed formulae for the induced field and current distribution, in the special case when $\kappa = k\rho^{-m}$, where k and m are constants are obtained and applied to the terrestrial magnetic variations The results obtained support the view, expressed by Chapman and Price, that there is a considerable increase of κ with increasing depth, beyond 150 km It seems however, that the really important increase in κ takes place at about 700 km depth beyond which κ is at least as great as 10^{-11} e m u while above this depth the mean conductivity may be of the same order as for rocks on the earth s surface (10^{-16} or 10^{-15} e m u) This suggests that there is some change in the composition of the earth (e g to a more metallic content) at a depth of about 700 km seismological evidence appears to indicate that such a transition occurs at a somewhat greater depth The results also show that there is an effective distribution of κ at or near the surface of the earth and it seems most probable that this represents the influence of the relatively highly conducting oceans The induced currents do not penetrate appreciably beyond a depth of about one fifth of the earth s radius, so that the knowledge of κ afforded by the daily variations and the storm time variations will be restricted to an outer shell of this thickness

---

The photochemical polymerization of methyl acrylate vapour

By H W Melville (*Communicated by E K Rideal F R S Received 28 March 1938*)

The photopolymerization of methyl acrylate vapour has been studied at pressures up to 60 mm and at temperatures between 20 and 150° C Polymerization occurs almost quantitatively giving a dense cloud of polymer in the gas phase when about 10 quanta are absorbed/c c /sec at 2537 A the quantum yield ranging from unity to 5000 depending on the temperature and intensity At shorter wave lengths the molecule is dissociated to hydrogen and propiolic ester Growth of the polymer ceases when two growing polymers interact with each other The temperature of this latter process is so large that the polymerization has a negative temperature coefficient

The reaction is very sensitive to inhibitors especially to oxygen and to a lesser extent butadiene By controlled use of inhibitors a direct measure of the energy of activation of the propagation reaction has been made it amounts to 4 k cal At high oxygen pressures oxygen exerts a positive catalytic effect believed to be due to the mercury sensitized production of O atoms

The molecular weight of the polymer has been estimated by measuring the ratio of acrylate to butadiene molecules used up in the inhibited reaction under a given intensity of illumination Hence by calculation the molecular weight of the product in the normal reaction is easily found

This behaviour of acrylate is in marked contrast with that exhibited by methyl methacrylate, the long life time of the latter polymer being absent in the acrylate Like the methacrylate the acrylate grows by the double bond mechanism and the differences in behaviour of these two molecules is accounted for by the theories advanced in this and previous papers

Indian Agricultural Research Institute (Pusa)
LIBRARY, NEW DELHI-110012

This book can be issued on or after ______________

Return Date	Return Date